MICROBIAL TRANSLOCATION IN CHILDREN WITH HIV IN UGANDA

THESIS SUBMITTED FOR COMPLETION OF A DOCTORATE OF PHILOSOPHY

FELICITY FITZGERALD

UNIVERSITY COLLEGE LONDON

INFECTION, IMMUNITY AND INFLAMMATION

UCL GREAT ORMOND STREET INSTITUTE OF CHILD HEALTH

30 GUILFORD STREET

LONDON

1 Declaration

I, Felicity Fitzgerald confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: 

Felicity Fitzgerald

Date: 9.7.17

2 Abstract

Background

There are >100000 new HIV infections in children each year, nearly 90% of which are in Africa. Untreated HIV results in activation across all axes of the immune system and this immune activation is linked to poor outcome. This study aimed to investigate microbial translocation (the crossing of microbial products such as bacterial DNA and components of bacterial cell walls from the gut into the blood stream) as a potential driver of immune activation and poor outcome in African children with HIV.

Method

The study included HIV-infected children initiating treatment with antiretroviral therapy (ART) in the clinical trial CHAPAS-3 in Uganda (ISRCTN69078957). ART naïve and ART-experienced children from urban (Kampala) and rural (Gulu) settings were included with age-matched HIV uninfected controls from urban communities. Plasma was collected at 3 time points from HIV infected children and a one-off plasma sample from HIV uninfected controls. Cell pellet samples from HIV-infected children in the urban setting and baseline were also available. Microbial translocation was assessed using broad range 16S rDNA polymerase chain reactions (PCR) with next generation sequencing (NGS) and a panel of specific bacterial PCRs. Intestinal barrier function was assessed using Intestinal fatty acid binding protein (I-FABP).

Results

In total, 305 children were included: 119 ART naïve (median age 2.8 years, interquartile range (IQR) 1.7-4, median baseline CD4% 20, IQR 14-24); 22 ART experienced children (median age 6.5 years IQR 5.9-9.2, median baseline CD4% 34, IQR 31-39) (urban site); 55 ART-naïve children (rural site)(median age 2 IQR 1.7-3.5, median baseline CD4% 30, IQR 22-41) and 109 age-matched HIV-uninfected controls.

For the HIV-infected groups, immune recovery was good. Most molecular assays including broad range PCR were negative or very low at all time points and across all groups. No relationship was seen between molecular assay results and clinical events including invasive infections in the HIV-infected group. I-FABP was significantly higher

in control groups than HIV infected (urban) groups at baseline, and the rural group was also higher. I-FABP increased in both urban HIV-infected groups over time.

Conclusion

In this longitudinal study, there was no convincing relationship between markers of microbial translocation and clinical progress in HIV infected children from Uganda after ART initiation. Levels of translocation were low and similar in HIV-uninfected controls. I-FABP may be a marker of a healthy gut in this setting. Microbial translocation does not play a pivotal role in driving immune activation in this setting.

3 Table of Contents

1 Declaration 3
2 Abstract 4
3 Table of Contents 6
4 List of Tables 10
5 List of Figures 12
6 Abbreviations 15
7 Acknowledgements 18
8 Chapter 8 Background 19
8.1 Human Immunodeficiency Virus (HIV) 19
8.2 The HIV Pandemic in Children: Current and Future Challenges 22
8.3 Immune Activation in HIV 27
8.3.1 Effects on the Immune System 29
8.3.2 End Organ Effects of Immune Activation 32
8.3.3 Causes of Immune Activation 33
8.4 Microbial Translocation in HIV 35
8.4.1 Mechanisms of Microbial Translocation 36
8.4.2 Evidence of Microbial Translocation in the Bloodstream 42
8.4.3 Microbial Translocation in Children with HIV 44
8.4.4 Interventions Targeting Immune Activation and Microbial Translocation 45
8.4.5 Challenges in Quantifying Microbial Translocation 45
8.4.6 Molecular Methods for Investigating Microbial Translocation in Blood 46
8.4.7 Methods for Assessment of Gastrointestinal Tract Injury 57
8.5 Pilot Study 57
8.6 Summary 59
8.7 Aims of the Thesis 60
8.7.1 Primary Hypothesis 61
8.7.2 Secondary Hypothesis 61
9 Chapter 9 Materials and Methods 62
9.1 Study Population 62
9.1.1 Sample Size Calculations 65
9.2 Ethical Considerations 65
9.3 DNA Extraction for Molecular Testing 66
9.4 Quantitative PCR 66
9.4.1 Identification of Target Organisms 66

9.4.2 Identification of Target Genes for Amplification 67
9.4.3 Designing Primer Probe Sets 67
9.4.4 Testing Primer Probe Sets on Samples with Known Bacterial Strains 67
9.4.5 Testing Experimental Samples using the Primer Probe Mixes 69
9.5 Broad Range 16S rDNA PCR 71
9.6 Sanger Sequencing 72
9.7 Next Generation Sequencing 72
9.8 Assessment of Gastrointestinal Tract Injury and Anthropometry 74
9.9 Comparison of Clinical Events to Molecular Test Results 75
9.10 Statistical Analysis 76
10 Chapter 10 Assay Development Results 77
10.1 QPCR Assay Development 77
10.1.1 Identifying Target Organisms 77
10.1.2 Identification of Target Genes for Amplification 79
10.1.3 Designing Primer and Probe Sets 80
10.1.4 Testing Primers against Target Strains 81
10.1.5 Testing Experimental Samples 82
10.2 Next Generation Sequencing Assay Development 83
10.2.1 Contamination of Barcoded Primers 83
10.2.2 Primer Dimer 84
10.3 Assay Development Discussion 90
10.3.1 Quantitative PCR Assay Development 90
10.3.2 NGS Assay Development 91
11 Chapter 11 Description of Study Cohort 94
11.1 Background 94
11.2 Study Population 94
11.3 Methods 94
11.4 Results 94
11.5 Discussion 102
12 Chapter 12 Quantitative PCR for the Detection of Specific Bacterial Species and Families as Indicators of Microbial Translocation 104
12.1 Background 104
12.2 Study Population 104
12.3 Methods 104
12.4 Results 104
12.4.1 Plasma Samples over Time 107
12.4.2 Pellet Samples 119

12.5 Discussion 122

13 Chapter 13 Broad Range 16S rDNA PCR for the Detection of Bacterial DNA as an Indicator of Microbial Translocation in HIV-infected and Uninfected Children. 127

13.1 Background 127

13.2 Study Population 127

13.3 Method 127

13.4 Results 128

13.4.1 Broad Range 16S rDNA PCR using SYBR® Green Assay 128

13.4.2 Comparison of Broad Range 16S rDNA PCR (using SYBR® Green assay) with qPCRs 134

13.5 Discussion 137

14 Chapter 14 Next Generation Sequencing subsequent to Broad Range 16S rDNA PCR to Identify Bacterial DNA involved in Microbial Translocation in HIV-infected and Uninfected children. 141

14.1 Background 141

14.2 Study Population 141

14.3 Methods 141

14.4 Results 142

14.4.1 Removal of OTUs in Negative Control Samples 147

14.4.2 Comparison of NGS Results with qPCR and Broad Range 16S rDNA PCR 151

14.5 Discussion 154

15 Chapter 15 Assessment of Gastrointestinal Tract Injury and Anthropometry in relation to Molecular Markers of Microbial Translocation in HIV infected and Uninfected children. 159

15.1 Introduction 159

15.2 Study Population 159

15.3 Methods 159

15.4 Results 159

15.4.1 I-FABP over time by ART Group 160

15.4.2 I-FABP compared with Broad Range 16S rDNA PCR (SYBR® Green) 165

15.4.3 Broad Range 16S rDNA PCR Results versus Anthropometric Measures 168

15.4.4 Comparison of S.aureus and Enterobacteriaceae qPCRs with Anthropometric Measurements 173

15.5 Discussion 176

16 Chapter 16 Clinical Events in HIV-infected Children during the Course of the Study in Relation to Molecular Markers of Microbial Translocation 180

16.1 Background 180

16.2 Study Population 180

16.3 Methods 180

16.4 Results 180
16.5 Discussion 192
17 Chapter 17 Discussion 195
17.1 Does Biologically Significant Microbial Translocation Occur in Children with HIV 195
17.2 Strengths and Limitations 204
17.2.1 Assay Limitations 204
17.2.2 Limitations in Availability of Samples 206
17.2.3 Cotrimoxazole 206
17.3 Future Research 207
18 References 210
19 Appendices 219
19.1 Appendix A Species in Microbial Mock Communities 220
19.2 Appendix B Tables of barcoded primer sequences for NGS 221
19.3 Appendix C QIIME script settings 224
19.4 Appendix D Missing samples across groups and time points 227
19.5 Appendix E Paper: "Evidence Microbial Translocation Occurs in HIV-Infected Children in the United Kingdom"; AIDS Research and Human Retroviruses; October 2013 228
19.6 Appendix F Conference Abstract "Microbial translocation does not drive immune activation in Ugandan children with HIV"; Conference for Retroviruses and Opportunistic Infections February 2017, Seattle, Washington; Abstract number 188 234

4 List of Tables

TABLE 1 *BIFIDOBACTERIUM* (*XFP*) PRIMER PROBE SET 67

TABLE 2 *LACTOBACILLUS* (*HSP60/GROEL* GENE) PRIMER PROBE SET 68

TABLE 3 *CLOSTRIDIALES* PRIMER PROBE SET 68

TABLE 4 REACTION CONSTITUENTS FOR QPCR REACTIONS 69

TABLE 5 CYCLING CONDITIONS FOR QPCR REACTIONS 71

TABLE 6 REAGENTS FOR NGS LIBRARY AMPLIFICATION 73

TABLE 7 CYCLING CONDITIONS FOR NGS LIBRARY AMPLIFICATION 73

TABLE 8 CUSTOM SEQUENCING PRIMERS 74

TABLE 9 SENSITIVITY OF ASSAY FOR DETECTING DIFFERENT *LACTOBACILLI* 82

TABLE 10 SCHEME FOR INTERPRETATION AND COMPARISON OF PCR RESULTS 83

TABLE 11 COMPARISON OF QUBIT AND KAPA RESULTS FOR LIBRARY QUANTIFICATION 85

TABLE 12 COMPARISON OF PRIMER DIMER TO AMPLICON CONCENTRATIONS AND QUBIT RESULTS FOR SECOND LIBRARY AMPLIFICATION 88

TABLE 13 TABLE DEMONSTRATING RELATIVE REDUCTION IN BOTH PRIMER DIMER AND AMPLICON CONCENTRATION USING REPEAT SIZE SELECTION WITH AMPURE BEADS 89

TABLE 14 COMPARISON OF RELATIVE PRIMER DIMER TO AMPLICON CONCENTRATIONS USING REDESIGNED PRIMERS AS ASSESSED BY AGILENT 2200 TAPESTATION 90

TABLE 15 DEMOGRAPHICS, CD4 T CELL COUNTS, VIRAL LOADS AND ANTHROPOMETRIC DATA COMPARED BY ART GROUP 96

TABLE 16 ART RANDOMISATION FOR HIV INFECTED CHILDREN 97

TABLE 17 AGE, ART RANDOMISATION AND CAUSES OF DEATH FOR 6 CHILDREN WHO DIED 102

TABLE 18 SAMPLE AVAILABILITY BY GROUP AT DIFFERENT TIME POINTS 105

TABLE 19 QPCR RESULTS FOR SAMPLES AT BASELINE COMPARING PROPORTIONS OF POSITIVE SAMPLES BY ART GROUP . 113

TABLE 20 WEEK 12 QPCR RESULTS COMPARING PROPORTIONS OF POSITIVE SAMPLES BETWEEN ART GROUPS. 114

TABLE 21 PROPORTIONS OF POSITIVE QPCR RESULTS AT WEEK 72 BY ART GROUP 115

TABLE 22 CHANGE IN PROPORTIONS POSITIVE FOR *S.AUREUS* OVER TIME BY ART GROUP 116

TABLE 23 TABLE COMPARING PROPORTIONS POSITIVE FOR *ENTEROBACTERIACEAE* OVER TIME BY ART GROUP 116

TABLE 24 TABLE COMPARING QPCR RESULTS FOR CELL PELLETS IN HIV-INFECTED CHILDREN AT BASELINE BY ART GROUP121

TABLE 25 BACTERIAL LOAD (CFU EQUIVALENTS) RESULTS BY BROAD RANGE 16S RDNA PCR ASSAY (SYBR® GREEN) BY ART GROUP OVER TIME COMPARED USING UNPAIRED WILCOXON RANK SUM TESTS 131

TABLE 26 PAIRED WILCOXON RANK SUM TESTS OF BACTERIAL LOAD (CFU EQUIVALENTS) OVER TIME BY ART GROUP ... 131

TABLE 27 SPEARMAN CORRELATION COEFFICIENTS BY ART GROUPS OVER TIME 133

TABLE 28 TABLE OF DISTRIBUTION OF SAMPLES SUCCESSFULLY SEQUENCED USING NGS BY ART GROUP AND SAMPLE TIME POINTS 142

TABLE 29 I-FABP BY ART GROUP OVER TIME (COMPARISONS USING UNPAIRED WILCOXON RANK SUM TEST) 162

TABLE 30 SPEARMAN CORRELATION COEFFICIENTS COMPARING I-FABP AT DIFFERENT TIME POINTS BY ART GROUP 165

TABLE 31 SPEARMAN CORRELATION COEFFICIENTS BETWEEN I-FABP AND BACTERIAL LOAD (CFU EQUIVALENTS BY BROAD RANGE 16S RDNA PCR) BY ART GROUP OVER TIME. 167

TABLE 32 CORRELATION COEFFICIENTS BETWEEN BACTERIAL LOAD (CFU EQUIVALENTS GENERATED BY BROAD RANGE 16S RDNA PCR USING SYBR® GREEN) AND ANTHROPOMETRIC MEASUREMENTS 171

TABLE 33 CORRELATION COEFFICIENTS BETWEEN *S.AUREUS* CODED NEGATIVE TO STRONG POSITIVE AND ANTHROPOMETRIC MEASUREMENTS AT BASELINE AND WEEK 72. 174

TABLE 34 CORRELATION COEFFICIENTS OF *ENTEROBACTERIACEAE* CODED NEGATIVE TO STRONG POSITIVE VERSUS ANTHROPOMETRY 175

TABLE 35 COMPARISON OF CLINICAL EVENTS OCCURRING WITHIN ONE MONTH OF A HIGH QPCR/16S RDNA PCR RESULT OR SUCCESSFUL NGS SEQUENCING 182

TABLE 36 DESCRIPTION OF CLINICAL EVENTS OCCURRING WITHIN ONE MONTH OF A POSITIVE TEST MOLECULAR TEST RESULT (ALL CHILDREN WERE ART NAÏVE AND FROM THE JCRC COHORT) 185

TABLE 37 ASSAY RESULTS FOR PATIENTS WITH POSITIVE BLOOD/URINE CULTURE 188

TABLE 38 SUMMARY OF PREVIOUS STUDIES INVESTIGATING MICROBIAL TRANSLOCATION IN THE BLOOD OF HUMANS INFECTED WITH HIV, INCLUDING METHODS USED FOR QUANTIFICATION AND STUDY RESULTS (AS REGARDS MICROBIAL TRANSLOCATION) 196

TABLE 39 INITIAL SET OF BARCODED PRIMERS 221

TABLE 40 SECOND SET OF BARCODED PRIMERS 223

5 List of Figures

FIGURE 1 STRUCTURE OF HIV REPRODUCED FROM SPLETTSTOESSER 2014 [3] 19

FIGURE 2 SCHEMATIC OF HIV LIFECYCLE REPRODUCED FROM SPLETTSTOESSER 2014[3] 21

FIGURE 3 SCHEMATIC OF HIV DISEASE PROGRESSION OVER TIME ADAPTED FROM PANTALEO *ET AL.* 1993[10] 22

FIGURE 4 NUMBER OF CHILDREN LIVING WITH HIV GLOBALLY (ADAPTED WITH PERMISSION FROM UNAIDS/WHO 2015)[13] 23

FIGURE 5 NUMBER OF CHILDREN RECEIVING ART (ADAPTED WITH PERMISSION FROM WHO 2016)[23] 24

FIGURE 6 SCHEMA OF INTERACTION BETWEEN HIV INFECTION, IMMUNE ACTIVATION AND MICROBIAL TRANSLOCATION (ADAPTED WITH PERMISSION FROM KLATT *ET AL.*) 28

FIGURE 7 SCHEMA OF IMMUNOSENESCENCE IN HIV (REPRODUCED FROM DEEKS 2011, NO PERMISSION REQUIRED)[65] 31

FIGURE 8 THE INTESTINAL EPITHELIUM IN HEALTH AND DURING HIV INFECTION, REPRODUCED WITH PERMISSION FROM SANDLER *ET AL.*, 2012[112] 38

FIGURE 9 SCHEMA OF BROAD RANGE 16S RDNA PCR..... 47

FIGURE 10 SCHEMATIC OF SANGER SEQUENCING TECHNIQUE (ADAPTED FROM WELLCOME TRUST SANGER INSTITUTE ANIMATION WITH PERMISSION)..... 50

FIGURE 11 SCHEMATIC OF ILLUMINA NEXT GENERATION SEQUENCING METHOD (ADAPTED FROM WELLCOME TRUST SANGER INSTITUTE ANIMATION WITH PERMISSION) 52

FIGURE 12 SCHEMA OF POSITIONING OF DUAL INDEX BAR CODES AND SEQUENCING PRIMERS FOR ILLUMINA MISEQ™ SEQUENCING METHOD (REPRODUCED WITH PERMISSION FROM ILLUMINA INDEXED SEQUENCING OVERVIEW GUIDE[181].53

FIGURE 13 SCHEMATIC OF QPCR METHODOLOGY ILLUSTRATING TAQMAN™ PRIMER PROBE METHOD..... 55

FIGURE 14 QPCR RESULTS SHOWING CYCLE THRESHOLD. 56

FIGURE 15 TRIAL PROFILE FROM CHAPAS 3 REPRODUCED UNDER CREATIVE COMMONS LICENSE FROM MULENGA *ET AL.* 2016[209] 63

FIGURE 16 AGILENT HSDNA KIT RESULTS DEMONSTRATING PRIMER DIMER IN FOUR PLATES OF LIBRARY PREPARATION..... 85

FIGURE 17 COMPARISON OF 0.8:1 (TOP PANEL) AND 0.6:1 (SECOND PANEL) AMPURE BEADS TO SAMPLE RATIO IN REDUCING PRIMER DIMERS IN PLATE NUMBERS 2, 3 AND 4 86

FIGURE 18 AMPLICON VERSUS PRIMER DIMER PEAKS USING AGILENT BIOANALYSER IN 4 PLATES OF POOLED LIBRARY POST-SIZE SELECTION AND ALTERING METHOD OF NORMALISATION 87

FIGURE 19 RESULTS OF AMPLIFICATION OF FOUR FURTHER 96 WELL PLATES LABELLED P1-P4 USING AGILENT BIOANALYSER DEMONSTRATING PRIMER DIMER AND AMPLICON PEAKS 88

FIGURE 20 BOXPLOTS OF CD4 T CELL COUNTS AND PERCENTAGES OVER TIME BY ART GROUP 98

FIGURE 21 HISTOGRAM OF VIRAL LOAD AT BASELINE AND SUPPRESSION (<100 C/ML) BY ART GROUP AT WEEK 96 99

FIGURE 22 BOXPLOTS OF ANTHROPOMETRIC MARKERS AT BASELINE BY ART GROUP 100

FIGURE 23 BOXPLOTS OF CHANGES IN ANTHROPOMETRIC MARKERS BY ART GROUP AT WEEK 96..... 101

FIGURE 24 BASELINE QPCR RESULTS RANKED NEGATIVE TO STRONG POSITIVE BY ART GROUP..... 109

FIGURE 25 WEEK 12 QPCR RESULTS RANKED NEGATIVE TO STRONG POSITIVE BY ART GROUP 110

FIGURE 26 WEEK 72 QPCR RESULTS RANKED NEGATIVE TO STRONG POSITIVE BY ART GROUP 111

FIGURE 27 CONTROL QPCR RESULTS RANKED NEGATIVE TO STRONG POSITIVE BY ART GROUP 112

FIGURE 28 CHANGE IN PROPORTIONS POSITIVE FOR *S.AUREUS* USING QPCR ASSAY OVER TIME WITH 95% CIS BY ART GROUP 117

FIGURE 29 CHANGE IN PROPORTIONS POSITIVE FOR *ENTEROBACTERIACEAE* USING QPCR ASSAY OVER TIME WITH 95% CIS BY ART GROUP 118

FIGURE 30 QPCR RESULTS AT BASELINE COMPARING PELLET WITH PLASMA SAMPLES BY ART GROUP: CFU EQUIVALENTS AS COMPARED WITH CT VALUES OF STANDARDS WITH KNOWN CFU QUANTITY AND CATEGORISED FROM NEGATIVE TO STRONG POSITIVE. 120

FIGURE 31 BOXPLOTS OF BACTERIAL LOAD (CFU EQUIVALENTS COMPARED WITH STANDARDS OF KNOWN CFUS) DETECTED BY BROAD RANGE 16S RDNA PCR (SYBR® GREEN) BY ART GROUP OVER TIME 130

FIGURE 32 CHANGE IN BACTERIAL LOAD (MEDIAN CFU EQUIVALENTS COMPARED WITH STANDARDS OF KNOWN CFUS) BY BROAD RANGE 16S RDNA PCR OVER TIME BY ART GROUP WITH IQR 132

FIGURE 33 COMPARISONS OF BACTERIAL LOAD (CFU EQUIVALENTS COMPARED WITH STANDARDS OF KNOWN CFUS) IN PELLETS VERSUS PLASMA SAMPLES AT BASELINE 133

FIGURE 34 SCATTERPLOTS OF QPCRS COMPARED WITH BROAD RANGE 16S RDNA PCR (HIGH NUMBER OF POSITIVE QPCRS), COMPARING RANKING OF NEGATIVE TO HIGH POSITIVES 135

FIGURE 35 COMPARISON OF QPCR VERSUS BROAD RANGE 16S RDNA PCR FOR PELLET SAMPLES, COMPARING RANKING OF NEGATIVE TO HIGH POSITIVES 136

FIGURE 36 PHYLOGENETIC TREE BY ART RANDOMISATION (UNRAREFIED): SHOWING PREDOMINANCE OF *STAPHYLOCOCCACEA, ENTEROBACTERIACEAE,* AND *SPHINGOMONADACEAE.* NEGATIVE EXPERIMENTAL CONTROL SAMPLES (COLOURED BLUE) CAN BE SEEN ACROSS MOST NODES 144

FIGURE 37 PHYLOGENETIC TREE BY ART RANDOMISATION (RAREFIED) SHOWING PREDOMINANCE OF *STAPHYLOCOCCACEA, ENTEROBACTERIACEAE,* AND *SPHINGOMONADACEAE.* NEGATIVE EXPERIMENTAL CONTROL SAMPLES (COLOURED BLUE) CAN BE SEEN ACROSS MOST NODES 145

FIGURE 38 PRINCIPAL COORDINATES ANALYSIS USING UNIFRAC (RAREFIED) SHOWING NO SPECIFIC CLUSTERING BY SAMPLE TYPE OR ART RANDOMISATION. NEGATIVE EXPERIMENTAL CONTROL SAMPLES (SQUARES) CAN BE SEEN DISPERSED ACROSS ANALYSIS WITH NO SPECIFIC CLUSTERING. 146

FIGURE 39 PHYLOGENETIC TREE BY ART RANDOMISATION AFTER REMOVAL OF NEGATIVE CONTROL OTUS (UNRAREFIED), SHOWING SPARSE OTUS INCLUDING *STAPHYLOCOCCACEAE* AMONGST OTHER FAMILIES. OTUS DERIVED FROM BOTH ART-EXPERIENCED AND ART-NAÏVE SAMPLES DISTRIBUTED ACROSS PHYLOGENETIC TREE. 148

FIGURE 40 PHYLOGENETIC TREE BY ART RANDOMISATION, NEGATIVE CONTROL OTUS REMOVED (RAREFIED) SHOWING SPARSE OTUS INCLUDING *STAPHYLOCOCCACEAE* AMONGST OTHER FAMILIES. OTUS DERIVED FROM BOTH ART-EXPERIENCED AND ART-NAIVE SAMPLES DISTRIBUTED ACROSS PHYLOGENETIC TREE. 149

FIGURE 41 PCOA AFTER REMOVAL OF NEGATIVE CONTROL OTUS (RAREFIED): NO CLUSTERING BY SAMPLE TYPE OR TIMING. ART NAÏVE AND ART EXPERIENCED SAMPLES DISTRIBUTED THROUGHOUT. 150

FIGURE 42 PCOA OF PELLET SAMPLES WITH NEGATIVE CONTROL OTUS REMOVED COMPARED WITH OTHER PCRS: BROAD RANGE 16S RDNA PCR, *ENTEROBACTERIACEAE* PCR AND *S. AUREUS* PCR. SOME CLUSTERING SEEN BY NEGATIVE AND HIGH RESULTS IN BROAD RANGE 16S RDNA PCR AND *S.AUREUS* PCR, BUT NOT BY NEGATIVE OR HIGH RESULTS IN *ENTEROBACTERIACEAE* PCR. 152

FIGURE 43 PCOA OF PLASMA SAMPLES AFTER NEGATIVE CONTROL OTUS REMOVED COMPARED BY OTHER PCR ASSAYS: BROAD RANGE 16S RDNA PCR, *ENTEROBACTERIACEAE* PCR AND *S. AUREUS* PCR. MINIMAL CLUSTERING SEEN BY NEGATIVE OR HIGH RESULTS IN ANY OF THE THREE ASSAYS. 153

FIGURE 44 BOXPLOTS OF I-FABP OVER TIME BY ART GROUP. 163

FIGURE 45 CHANGE IN MEAN I-FABP OVER TIME BY ART GROUP WITH 95% CONFIDENCE INTERVALS 164

FIGURE 46 SCATTERPLOT COMPARISONS OF I-FABP AND BACTERIAL LOAD (BROAD RANGE 16S RDNA PCR USING SYBR® GREEN) BY ART GROUP AT DIFFERENT TIME POINTS 166

FIGURE 47 SCATTERPLOT OF BACTERIAL LOAD (CFU EQUIVALENTS GENERATED BY BROAD RANGE 16S RDNA PCR USING SYBR® GREEN) COMPARED TO ANTHROPOMETRIC MEASUREMENTS AT BASELINE 169

FIGURE 48 SCATTERPLOT OF WEEK 72 BACTERIAL LOAD (CFU EQUIVALENTS GENERATED BY BROAD RANGE 16S RDNA PCR USING SYBR® GREEN) COMPARED TO WEEK 96 ANTHROPOMETRIC MEASUREMENTS 170

FIGURE 49 PCOA OF RAREFIED SAMPLES, BY CLINICAL EVENT/NO EVENT USING UNIFRAC SHOWING NO CLUSTERING ACCORDING TO WHETHER OR NOT A CLINICAL EVENT OCCURRED. 190

FIGURE 50 PCOA OF RAREFIED SAMPLES (UNIFRAC), NEGATIVE CONTROL OTUS REMOVED BY PRESENCE/ABSENCE OF CLINICAL EVENT, SHOWING NO CLUSTERING BY WHETHER OR NOT A CLINICAL EVENT OCCURRED 191

6 Abbreviations

95% CI	95 per cent Confidence Interval
ABC	abacavir
AHR	Adjusted hazard ratio
AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
AZT	zidovudine
BLASTN	Nucleotide basic local alignment search tool
BMI	Body mass index
Bp	Base pair
CFU	Colony forming unit
CI	Confidence Interval
CMV	Cytomegalovirus
CT	Cycle threshold
D4T	stavudine
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
EFV	Efavirenz
EndoCab	Anti-endotoxin core antibodies
FASTA	Fast-All format
FASTQ	Fast-Quality format
FLASH	Fast length adjustment of short reads
GALT	Gut associated lymphoid tissue
GOSH	Great Ormond Street Hospital

HAART	Highly active antiretroviral therapy, also referred to as ART.
HAD	HIV-associated dementia
HAZ	height-for-age Z-score
HIV	Human immunodeficiency virus
hsCRP	High sensitivity C-reactive protein
IDO	Indoleamine 2,3-dioxygenase-1
I-FABP	Intestinal fatty acid binding protein
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
iNKT	Invariant natural killer T cells
JCRC	Joint Clinical Research Centre, Kampala
KT	Kynurenine to tryptophan
LBP	Lipopolysaccharide-binding protein
LPS	Lipopolysaccharide
mDCs	myeloid dendritic cells
MRC CTU	Medical Research Council Clinical Trials Unit
MT	Microbial translocation
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NK	Natural Killer
NHL	Non-Hodgkins lymphoma
NNRTI	Non-nuceloside reverse transcriptase inhibitor
NVP	Nevirapine
OTU	Operational taxonomic unit
PAMP	Pathogen associated molecular pattern

PMTCT	Preventing mother to child transmisssion
PCR	Polymerase chain reaction
PCoA	Principal Coordinates Analysis
QIIME	Quantitative insights into microbial ecology
qPCR	Quantitative polymerase chain reaction
rDNA	Ribosomal deoxynucleic acid
RM	Rhesus macaques
sCD14	Soluble CD14
SD	Standard deviation
SDG	Sustainable Development Goal
SIV	Simian immunodeficiency virus
SM	Sooty mangabey
Th17	T-helper 17
TLR4	Tol-like receptor 4
Tm	Melting temperature
TNFα	Tumour necrosis factor alpha
Tregs	T regulatory cells
WAZ	Weight-for-age z score
WHO	World Health organisation
WHZ	weight-for-height Z-score
µL	microlitre

7 Acknowledgements

I am very grateful to my supervisors, Nigel Klein, Kathryn Harris, Andrew Prendergast and Diana Gibb for their constant support and encouragement throughout this project, and to Sarah Walker for her incisive advice, both statistical and otherwise. I have also been well supported by the CHAPAS-3 Trial Team: in particular Cissy Kityo, Julia Kenny, Adrian Cook, Victor Musiime, George Abongomera, and the very patient and supportive Ellen Owen Powell. In the laboratory, I have had ceaseless support and encouragement from Ronan Doyle, Julianne Brown, Simon Eaton and Tony Brooks, and Liam Shaw has guided me through R Studio. Huge thanks go also to Vania Toledo without whom the department would not function. Of course, I could not have completed this thesis without the support of my family and friends: the biggest thanks go to them.

8 Chapter 8 Background

8.1 Human Immunodeficiency Virus (HIV)

The human immunodeficiency virus (HIV) is an RNA lentivirus within the *Retroviridae* family which can cause acquired immune deficiency syndrome (AIDS) in humans[1]. It consists of two copies of positive single stranded RNA within a capsid which also contains vital enzymes for viral replication: integrases, proteases and reverse transcriptases, and transcriptional transactivator (Tat) proteins(Figure 1). The virus envelope is a lipid bilayer formed from the host cell membrane expressing the HIV envelope protein Env, consisting of the gp120 and gp41 glycoproteins[2].

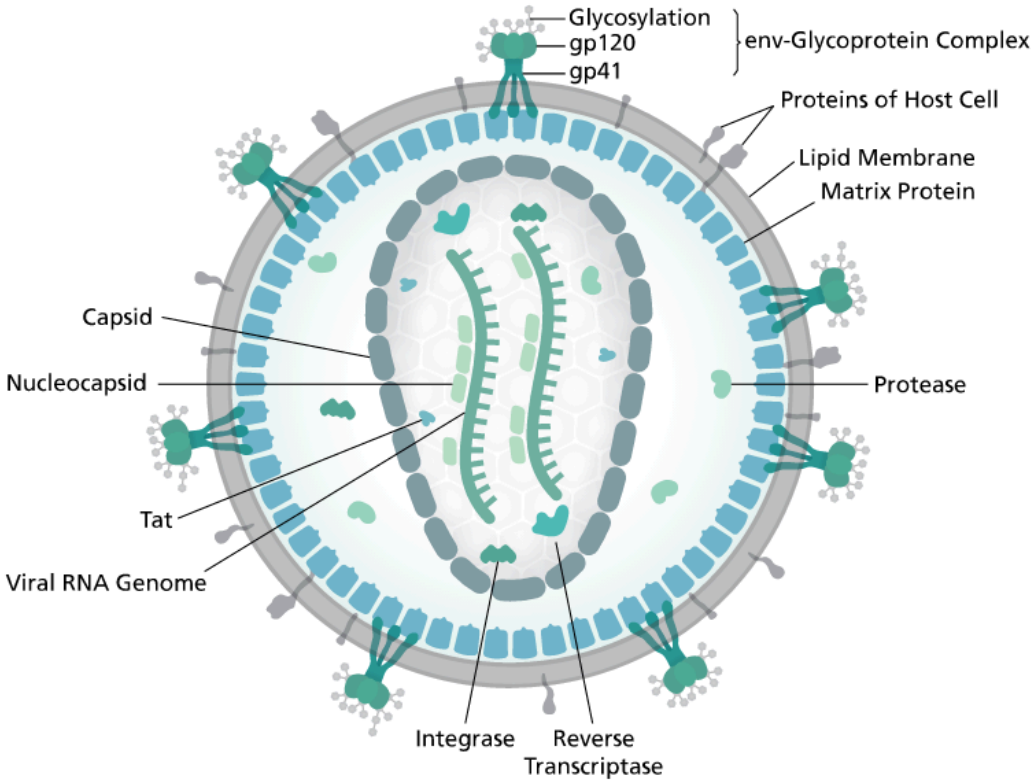


Figure 1 Structure of HIV reproduced from Splettstoesser 2014 [3]

Schematic of HIV viral structure, showing envelope proteins gp120 and gp41, the positions of integrase and protease enzymes, the viral RNA within the capsid and the Tat (transcriptional transactivator) proteins. Reproduced under Creative Commons Licence.

The virus is tropic for the cluster differentiation 4 (CD4) antigen found on T helper cells, dendritic cells and macrophages. The viral protein gp120 binds with host cell

CD4 molecules, and viral adsorption is triggered with concomitant viral interaction with host CXCR4 and CCR5 co-receptors (Figure 2)[4,5]. This is followed by fusion of viral and host cell membranes, releasing the viral capsid into the host cell. The viral reverse transcriptase then copies viral RNA into double stranded complementary DNA (cDNA) in an error-prone process which is one of the main drivers of viral mutation[6]. The viral DNA is integrated into host DNA in the nucleus using the viral integrase enzyme. At this point the virus can enter a latent phase, particularly in macrophages which can act as long standing reservoirs for HIV[7]. Viral replication takes place in the presence of host transcriptional factors such as the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) which is upregulated in activated T-cells[1]. This triggers transcription of the integrated proviral DNA into RNA, some of which is translated into viral proteins such as Tat which further upregulates viral production and the structural proteins Gag and Env[8]. The virion is assembled, and buds out of the host cell with the protease enzyme mediating a crucial cleavage of the gag polyprotein into matrix, capsid and nucleocapsid proteins[1].

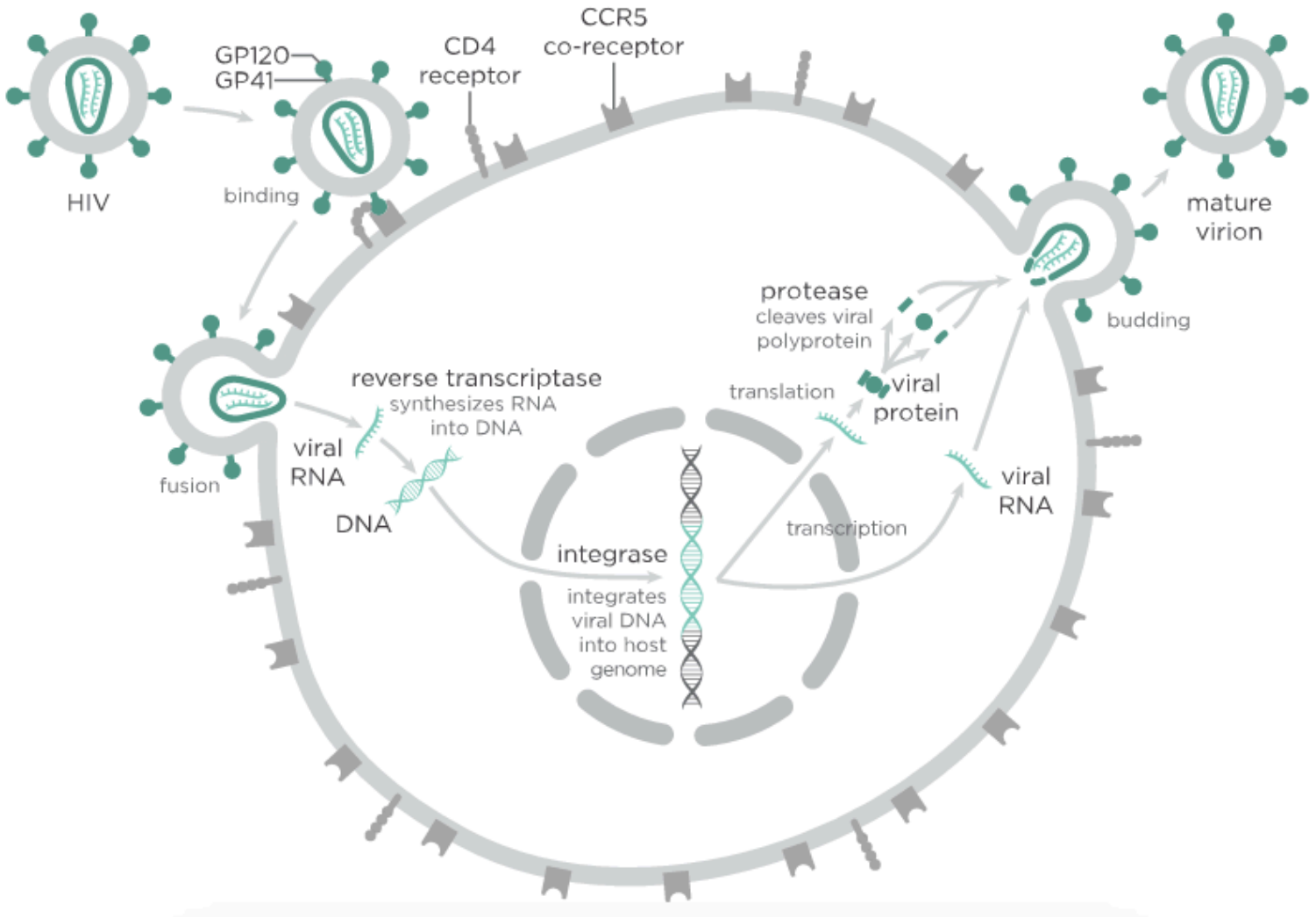


Figure 2 Schematic of HIV lifecycle reproduced from Splettstoesser 2014[3]

Schematic of HIV life cycle showing membrane proteins gp120 and gp41 binding with CD4 and CCR5 receptors on CD4 T cell prior to membrane fusion, reverse transcription and integration of viral DNA into host DNA. This is followed by transcription of the viral RNA and translation of viral proteins, before cleavage of viral proteins by a protease and assembly of the virion. The final stage is budding of the mature virus from the host cell. Reproduced under Creative Commons Licence.

The virus is then released to infect further CD4 T cells, resulting over time in a dramatic depletion of CD4 T cells and decline in cell-mediated immunity with concomitant increase in viral load (Figure 3). This depletion of cell-mediated immunity over time leaves HIV-infected people vulnerable to opportunistic infections that were the hallmark of AIDS in the pre-ART era. Since the advent of ART, AIDS is no longer an inevitably terminal state in that those attending hospital with severe opportunistic infections or AIDS-related malignancies (previously known as "AIDS-defining" conditions), can reconstitute their immune system on initiation of therapy, although mortality remains increased. Disease progression has been defined according both to the level of CD4 T-cells and the presence of specific opportunistic infections that occur with diminishing immunity[9]. Thus the disease is classified into four clinical stages by the World Health Organization, with WHO Stage 1 being asymptomatic and Stage 4 being the most severe[9].

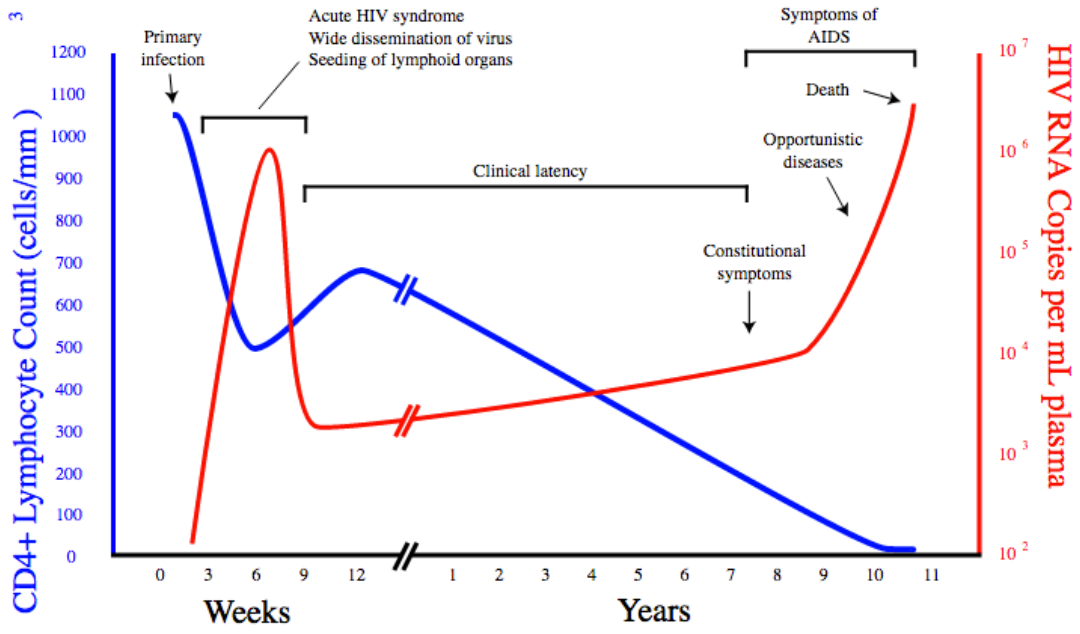


Figure 3 Schematic of HIV disease progression over time adapted from Pantaleo *et al.* 1993[10]

Schematic graph of relationship between CD4 T-cell count and HIV copies (viral load) during course of untreated HIV infection: individual time courses may vary. Blue line: CD4 T cell count (cells/μL). Red line: HIV RNA copies per mL plasma. Reproduced with permission from Pantaleo *et al.* Copyright Massachusetts Medical Society [10]

8.2 The HIV Pandemic in Children: Current and Future Challenges

In 2015 an estimated 36.7 million people were living with HIV globally, including 1.8 million children <15 years of age[11]. Significant strides have been made towards the United Nations Sustainable Development Goal (SDG) of reducing new HIV infection to zero by 2030, but there is still much work to be done. Although the rate of new infections in children has fallen from a peak of 490000 per annum in 2000 there were still 110000 new paediatric infections in 2015[12]. Nearly 90% of these HIV-infected children are in Africa (Figure 4)[13].

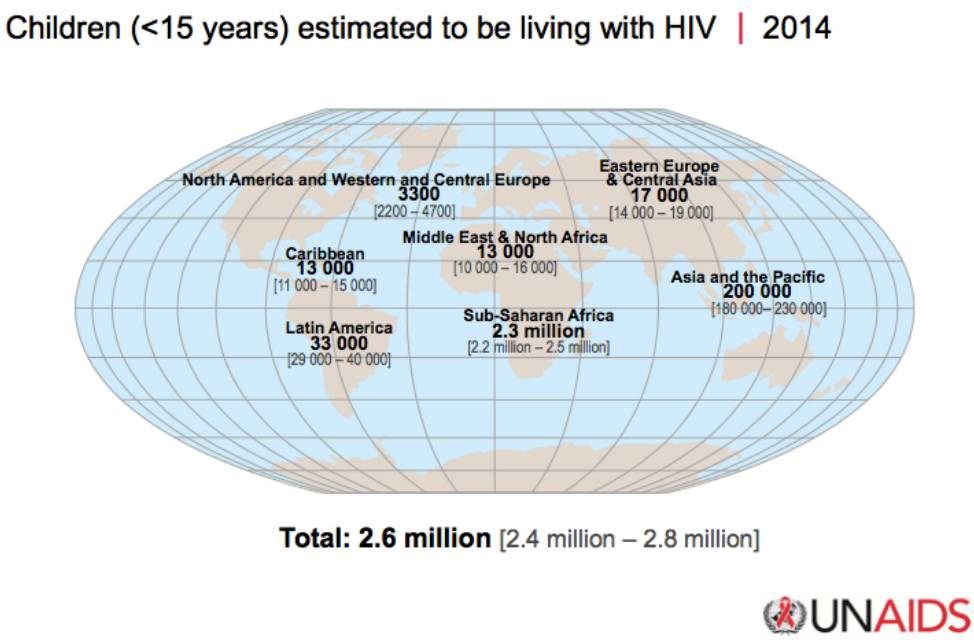


Figure 4 Number of Children Living with HIV globally (adapted with permission from UNAIDS/WHO 2015)[13]

ART, combined therapy usually with three antiretroviral agents, has revolutionised the management of HIV. It has dramatically increased the life expectancy of both adults and children with HIV, as well as reducing both horizontal transmission between serodiscordant partners and vertical transmission in prevention of mother-to-child transmission (PMTCT) programmes[14,15]. Mortality rates in HIV infected children in African settings were over 50% by age 2 prior to widespread antiretroviral therapy (ART) roll out[16]. In a cohort of >3500 children with HIV in the United States, death rates fell from 7.2 per 100 person years in 1994 to 0.8 from 2000-2006[17]. Similarly in the United Kingdom, the combined rate of AIDS defining events and mortality declined from 13.3 cases per person years before 1997 to 2.5 in the years

from 2003-2006[18]. In low/middle income settings, mortality rates and rates of disease progression have also decreased with ART roll out: in a large clinical trial (the ARROW Trial) in Uganda and Zimbabwe, rates of death or WHO Stage 4 events combined was 1.7-2 per 100 child years, and mortality rates were 1.1-1.3 per 100 child-years[19]. National ART programmes have also seen improved survival rates to adolescence in Sub-Saharan Africa[20-22].

Herculean efforts have been made to roll out ART globally: over 15.8 million people are now estimated to now be receiving treatment, with a concomitant (although slower) increase in the number of children accessing ART to just under 900000 (Figure 5)[12,23]. On the basis of the South African CHER trial which demonstrated that early ART reduced mortality fourfold in infants compared with deferred therapy, together with other observational data, the World Health Organization (WHO) guidelines were altered in 2008 to advise commencing all HIV-infected children under 2 on ART[24,25]. In 2016, based on data from the adult START trial, this recommendation was extended to all people living with HIV in 2016[9].

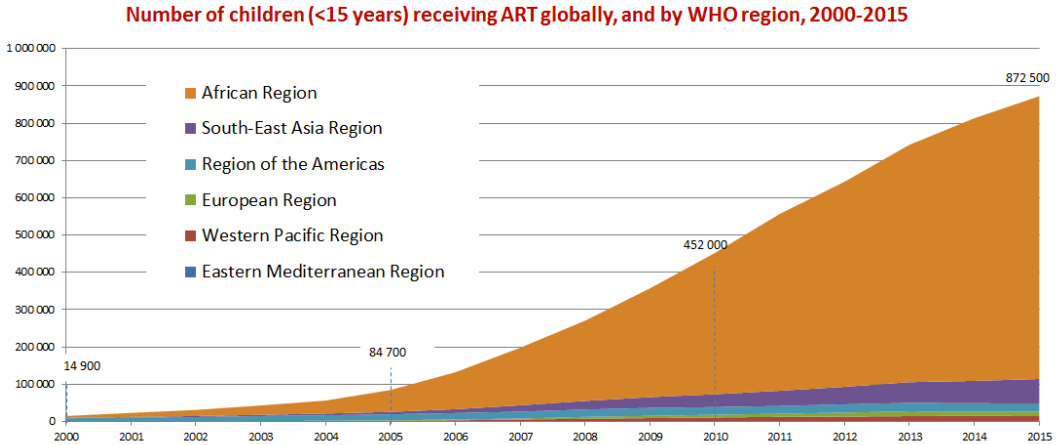


Figure 5 Number of children receiving ART (adapted with permission from WHO 2016)[23]

PMTCT programmes play a crucial role in the reduction of paediatric HIV. These programmes were among the first to use ART for prevention of ongoing HIV transmission. In 2012, on the basis of trial and programmatic data the WHO recommended the "Option B+" for pregnant women, whereby all women newly diagnosed with HIV commence ART during pregnancy and remain on it for life, rather than stopping therapy after the time of potential exposure for the infant[26]. This simplification of regimens had been shown to be effective in reducing numbers of infected infants Malawi in field settings[27]. As most of the 22 high priority countries have now adopted "Option B+", and in line with the revised recommendation of ART initiation for all, the WHO has further simplified guidelines to remove other options and advocate that all pregnant women should start and continue triple therapy with ART[9]. In terms of the efficacy of PMTCT, transmission rates of as low as 1% were achieved in a trial setting in Botswana, and rates of <1% transmission are now seen routinely in high income settings[15,28].

However, despite ART scale up, less than half of children living with HIV were estimated to be accessing treatment in 2015, and there were approximately 150000 deaths in children <15 years due to HIV[12]. Delays in early infant diagnosis and losses to follow up meant that in 2014 only 50% of perinatally infected infants were tested in the first two months of life, and in previous years only 30% of infants completed the cascade from diagnosis to ART initiation[29]. There is increasing evidence that early ART initiation promotes better long term immune reconstitution in children, and that even without evidence of immune suppression, there may be other pathways leading to increased morbidity and mortality (namely that of inflammation)[30,31]. Streamlining the early infant diagnosis cascade by effective identification of high-risk infants, minimising delays between testing and mother-infant pairs receiving test results, and starting appropriate treatment is a crucial priority to reduce both early infant deaths and later morbidity in children on ART.

Even once initiated on ART, the long-term health implications for children living with HIV are substantial. Mortality in the first year of ART can be up to 20% with

sepsis/meningitis and pneumonia being common causes of death[32,33]. However, in general, mortality is low if ART is started promptly, with many children surviving to adolescence, both in high income and low/middle income countries, and living with HIV as a chronic condition[20,21,34]. This changing population brings its own challenges. During the turbulent time of adolescence, adherence can drop, with a high risk of loss to follow up[35,36]. Triple-class virological failure was found in 10% of a northern cohort of 1007 adolescents with perinatal HIV infection, significantly associated with delayed treatment initiation and with longer time on ART[37]. The implications of virological failure on future treatment options for these children is still unclear but likely to represent significant challenges for adult services caring for vertically-infected young people both in low and high income settings[34,38].

Long term chronic morbidity for children and adolescents living with HIV is a growing concern. For children with HIV who have survived to adolescence without ART, these conditions included poor growth, recurrent respiratory infections, diarrhoea and chronic lung disease as well as more classic acute opportunistic infections[38,39]. Stunting, with poorer catch up after ART commencement compared to ponderal growth, appears to affect those in African settings more than European. This may reflect higher background rates of malnutrition[40]. Neurocognitive dysfunction, skin diseases, renal and bone disease are recognised features of adolescent HIV[38]. There is also the question of increased future risk of non-AIDS related conditions such as malignancy and cardiovascular disease. In terms of cardiovascular risk, a large study of HIV-infected children enrolled in the CHAPAS 3 trial in Uganda and Zambia with some overlap with the population in this study (see section 9.1), demonstrated significantly poorer markers of arterial stiffness and carotid intima media thickness in HIV-infected children than age-matched controls from the same community[41]. Perinatally infected children and adolescents with HIV in the United States had high aggregate cardiovascular risk scores when assessed using the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) scoring system[42]. Indeed, in both high income and low/middle income settings, cardiac pathologies such as ventricular hypertrophy are a feature of perinatally infected adolescents[43,44].

In summary, ART roll out has dramatically increased the longevity of people living with HIV, including children. However there are still significant challenges to be faced in the field of paediatric HIV. First, further reducing the numbers of infants vertically infected; second, addressing gaps in the cascade of early infant diagnosis, to ensure all HIV-infected infants access treatment promptly; third, ensuring ART regimens and associated therapy minimise development of acute and chronic complications of HIV, including ART toxicity; and finally ensuring adequate viral suppression to preserve future treatment options for children and adolescents needing lifelong ART. This thesis addresses the third challenge: investigating microbial translocation and immune activation as potential drivers of chronic complications in children with HIV.

8.3 Immune Activation in HIV

Untreated HIV results in a state of chronic immune activation across all the axes of the immune system. Both CD4 and CD8 T-cells are affected, with increased expression of CD38 and HLA-DR along with increased expression of the death receptor Fas (CD95) on CD8 cells[45]. Indeed, higher levels of activated CD8+CD38+CD45RO+ cells predicted CD4 T cell decline in the pre-ART era better than viral load, as did increased expression of CD38 on CD4 T cells in adults[46,47]. This immune activation and consequent poorer CD4 T cell count recovery are linked to increased rates of non-AIDS related diseases and increased all-cause mortality in patients on ART[48-53] (Figure 6).

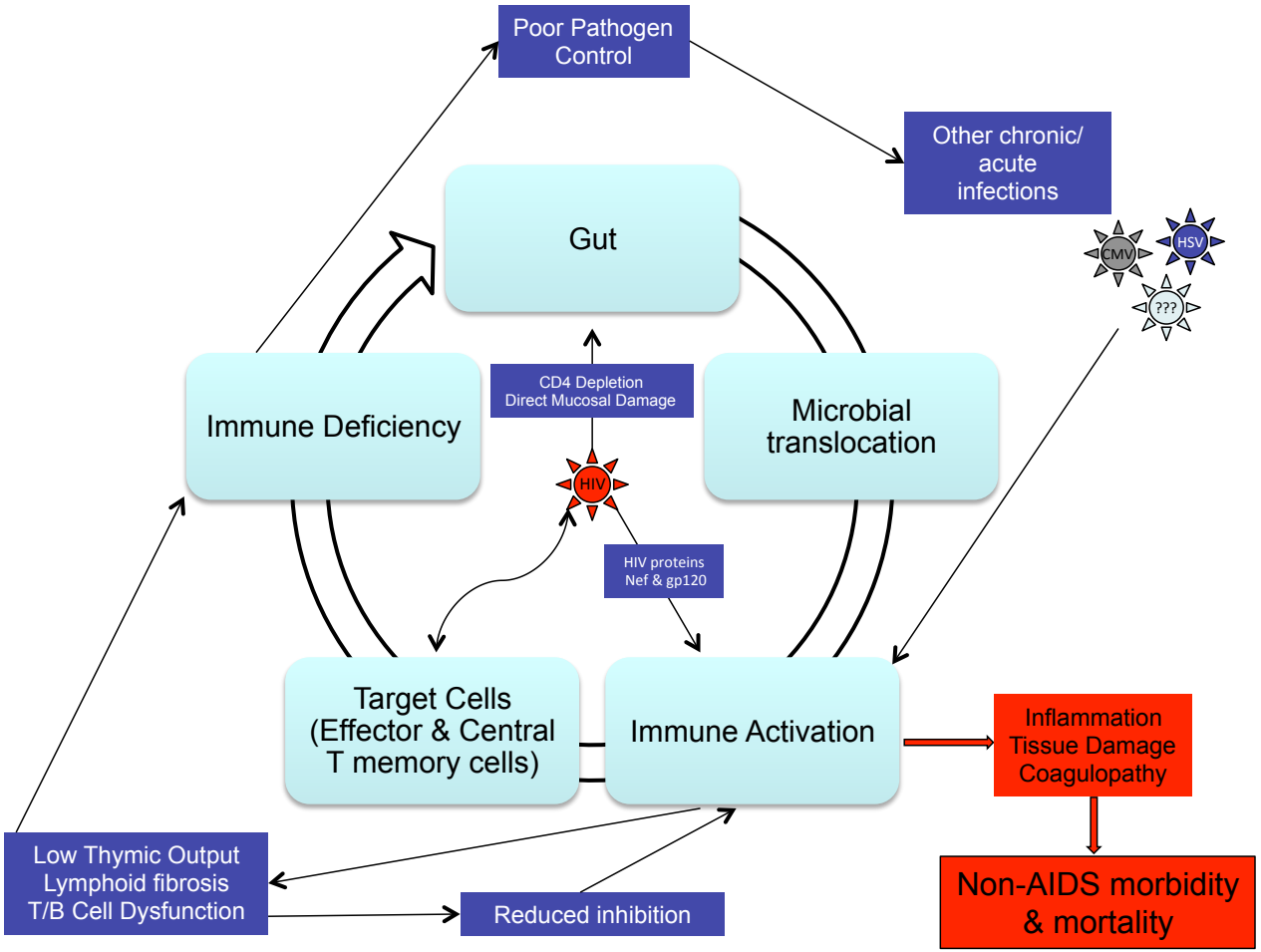


Figure 6 Schema of interaction between HIV infection, immune activation and microbial translocation (adapted with permission from Klatt *et al*.)

HIV initiates and sustains a cycle of immune activation. Demonstrating hypothesised interaction of i) direct immune activation by HIV proteins ii) Reduced inhibition by dysfunctional immune system iii) other chronic and acute infections such as cytomegalovirus (CMV) or herpes simplex virus (HSV) iv) microbial translocation as mechanisms of immune activation leading to non-AIDS pathogenesis. Adapted with permission from Klatt *et al*.[54]

Monocytes and dendritic cells are also activated. An early study found an expansion of the monocyte $CD14^{low}CD16^{high}$ subset expressing HLA-DR in untreated HIV, with an associated increase in expression of pro-inflammatory cytokines Interleukin-1 (IL-1), and Tumour necrosis factor alpha (TNFα)[55]. The enzyme indoleamine 2,3-dioxygenase-1 (IDO) is expressed in activated monocytes and dendritic cells and produces kynurenine and downstream toxic metabolites of tryptophan. In Ugandan adults with HIV initiating therapy, kynurenine produced by these activated monocytes and dendritic cells independently predicted poor CD4 T cell recovery and increased mortality[56]. Alongside increased production of IL-1, TNFα and IL-6 there is polyclonal B cell activation and resultant hypergammaglobulinaemia [57-59]. Indeed B cell activation and dysfunction was one of the earliest described hallmarks of AIDS[60]. Consequences of immune activation may be divided into the impact on the immune system itself and on end organs.

8.3.1 Effects on the Immune System

Effects on the immune system include the immediate consequence both of generating a larger pool of activated CD4 T cells expressing CCR5 and so vulnerable to HIV infection, and the upregulation of NF-κB which in turn activates transcription of HIV itself and thus increases viral replication[61]. However, the proportion of CD4 T cells infected with HIV (0.01-1%) during chronic infection is too low to account for the decline in CD4 T cell count seen in HIV progression- a discrepancy which may be accounted for by a further consequence of immune activation, that of immunosenescence[62] (Figure 7). Immunosenescence refers to the normal process of aging of the immune system seen in older adults. It involves involution of the thymus, a decrease in haematopoietic stem cells, reduced naïve T-cells with concomitant increase in well-differentiated memory T-cells with limited capacity to proliferate, and increased levels of proinflammatory cytokines[63-65]. In HIV, it appears that this process is accelerated. Through multiple pathways including bystander apoptosis, pyroptosis mediated by caspase-1, thymic involution, lymphatic fibrosis and clonal exhaustion, there is a gradual decline in the ability of the immune system to renew

naïve T cell pools, produce HIV-specific T cells and control HIV infection (Figure 7)[66-71]. In particular, caspase-1-mediated pyroptosis of CD4 T cells due to abortive infection with HIV is intensely inflammatory and could provide a key link between immune activation and the dramatic CD4 T cell depletion seen in HIV[68].

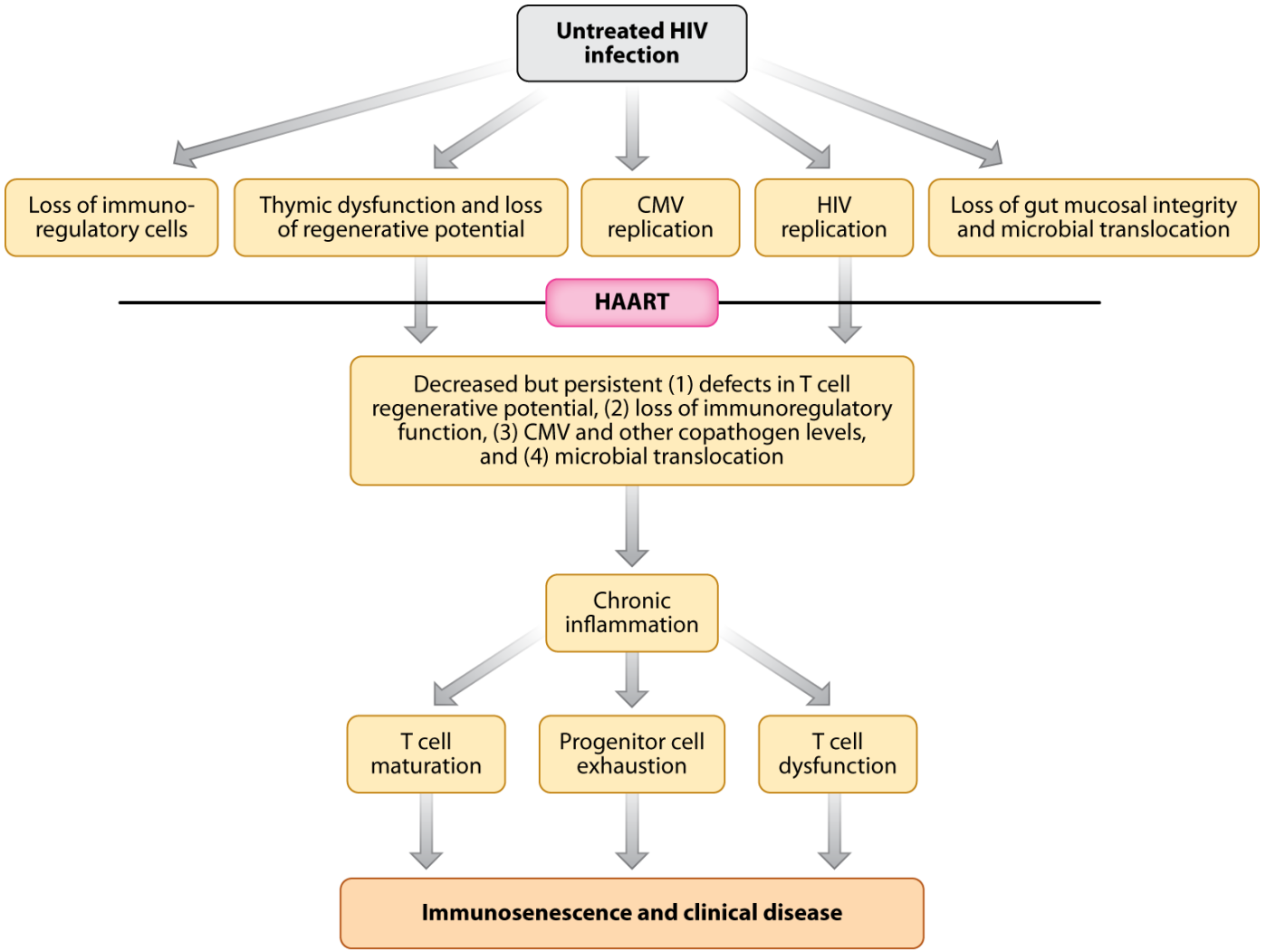


Figure 7 Schema of immunosenescence in HIV (reproduced from Deeks 2011, no permission required)[65]

HARRT: Highly active retroviral therapy or ART.

Children with HIV may be more resilient to this process of immunosenescence, with an increased thymic output acting as the source for a restored CD4 T cell population after commencing ART, dependent on the initial stage of CD4 T cell depletion[72]. However, when comparing children on ART who achieve viral suppression with those who do not, those who fail to achieve suppression had persistently higher levels of activated CD8+CD38+ T cells versus those who achieved suppression, alongside higher lipopolysaccharide (LPS: a component of Gram negative bacterial cell walls, see section 8.4), and lower numbers of central memory, effector/memory and CD38+ T cells[73].

Support for the importance of immune activation in the pathogenesis of HIV is seen in simian models of SIV infection. Sooty mangabeys are natural hosts of SIV and maintain normal CD4 T cell counts and remain clinically well despite high viral loads. They do not demonstrate immune activation. By contrast, rhesus macaques (RMs) which are not natural hosts of SIV, have aberrant immune activation with subsequent immune exhaustion and experience a similar progressive immunodeficiency syndrome to humans with HIV[74].

8.3.2 End Organ Effects of Immune Activation

Markers of immune activation have been consistently linked to increased all-cause mortality and disease progression in HIV (Figure 6)[46,53,75]. In the SMART study, high levels of IL-6, high sensitivity C-reactive protein (hsCRP) and D-Dimer were all found to be associated independently with an increased risk of mortality from non-AIDS events, particularly cardiovascular disease. In co-infection of HIV with either hepatitis B or C, increased TNFα was associated with a more rapid rate of liver fibrosis, compared with patients with lower levels of TNFα[76]. Plasma soluble CD14, a marker of monocyte activation was found to be associated with poorer scores in cognitive function tests indicating global impairment and with a diagnosis of HIV-associated dementia (HAD), as well as increased cardiovascular risk[52,77,78]. Kynurenine production from IDO-expressing activated monocytes has also been linked to HAD[79].

In pathogenic SIV, immunohistochemical staining for LPS and *Escherichia coli* in the liver was associated with the recruitment of hypercytotoxic Natural Killer (NK) cells and subsequent markers of liver damage, hepatitis and fibrosis at necropsy[80]. Increased B cell activation appears to be implicated in the development of AIDS-Non-Hodgkins lymphoma (AIDS-NHL) of the B cell lineage, and finally raised renal interstitial and glomerular non-polymorphic MHC Class II, interferon (IFN)-alpha and IFN-gamma receptor protein were found to be positively associated with HIV associated nephropathy[51,81].

8.3.3 Causes of Immune Activation

Immune activation in HIV-infected individuals is likely to be multifactorial with 4 main pathways implicated: i) driven directly by HIV proteins ii) immune response to HIV infection including loss of regulatory T-cells iii) re-activation or primary infection with other persistent infections such as cytomegalovirus (CMV), Epstein-Barr virus (EBV) and tuberculosis; and iv) driven by translocation of microbial products from the gut (Figure 6).

i. HIV proteins such as gp120 and Nef can cause immune activation not only by causing direct infection of T cells but also indirectly. For example, gp120/chemokine interactions on macrophages lead to release of pro-inflammatory cytokines such as TNFα and macrophage-inflammatory protein 1β[82]. The Nef protein, amongst other roles, appears to be responsible for the down-regulation of the negative T cell modulator CTLA-4 so leading to uncontrolled T cell activation.[83]

ii. Increasingly, the process of maladaptive immune activation appears in part to be due to reduced inhibition by regulatory mechanisms, such as the interaction between Nef and CTLA-4 described above. In a study of untreated HIV infected individuals, a significant negative relationship was found between declining numbers of T regulatory cells (Tregs) and CD8 activation (HLA-DR and CD38) ($R = -0.33$, $P = 0.03$)[84]. This study supports other work which suggest that Tregs have a protective role in limiting immune activation,

and their decline is associated with an increase in immune activation both in HIV and simian immunodeficiency virus (SIV) infection[85-87]. In a virus-free mouse model, activated CD4 T-cell ablation (to mimic the destruction of these cells by HIV) resulted in generalised immune activation which was prevented by reconstitution of regulatory T cells.[88]

iii. Other infections, either chronic and reactivating or as acute primary infections may also have a role. CMV in particular appears to be associated with poor outcomes in HIV infection: a cohort study of over 6000 Italian HIV infected individuals found that those with a positive CMV-Immunoglobulin (Ig) G at baseline were more likely to suffer severe non-AIDS events or death (adjusted hazard ratio (AHR) 1.53, 95% confidence intervals (CI):1.08-2.16) with a non-significant trend towards increased risk of cardiovascular disease[89]. This supports previous evidence that in HIV infected individuals, CMV specific T cell responses and T cell activation were both independently associated with increased carotid intima media thickness (a precursor to atherosclerosis)[90]. A further study found significantly increased levels activated T cells specific for herpes viruses (persistent infections) when compared to HIV negative controls and also to the non-persistent antigen tetanus toxoid[91]. In Ugandan HIV infected adults, co-infections were significantly associated with CD4 T cell activation[92]. These infections included tuberculosis (40%), candida (30%), malaria (13%), and herpes zoster (13%)[92]. Visceral leishmaniasis has also been implicated[93].

iv. Microbial translocation as a cause of immune activation in HIV is the principal focus of this thesis and is discussed at greater length below.

8.4 Microbial Translocation in HIV

It is postulated that rapid depletion of CD4 T cells within gut-associated lymphoid tissue (GALT) in early HIV infection allows microbial translocation from the gut[94-97]. Bacterial products have been identified at higher levels in the bloodstream of patients with HIV as compared to uninfected controls[98,99]. These products include potential immunostimulants, including lipopolysaccharide (LPS), a component of Gram-negative bacteria, and 16S ribosomal DNA (16S rDNA), common to all bacterial species. Increased circulating LPS is associated with higher levels of immune activation in HIV-infected patients[52,99]. Higher levels of 16S rDNA detected with broad-range polymerase chain reactions (PCR) and LPS were associated with poorer immune restoration in patients on ART[98]. Immune activation in those with poor CD4 T cell responses to ART has been associated with *Enterobacteriaceae* detected in the blood by 16S rDNA PCR and sequencing[100].

Microbial translocation as a mechanism for inducing immune activation was first proposed by Brenchley *et al.* in 2006.[99] Having noted the dramatic decline of CD4 T cells in the gut associated lymphoid tissue (GALT) early in HIV infection demonstrated by Veazey *et al.*[94] and their own earlier work[97], they quantified LPS in monkeys with pathogenic and non-pathogenic SIV, finding a significant difference between hosts and over the course of disease (see Figure 8 and section 8.4.1 for description of HIV-related gut damage). In the pathogenic model they found evidence of the LPS being bio-active, in that there was a significant increase in the level of soluble CD14 (sCD14: a marker of monocyte activation produced particularly in response to LPS) with increasing LPS. *In vitro*, it was possible to induce CD8 activation in HIV negative cells by culturing with plasma from HIV infected individuals containing increased levels of LPS. In a cross sectional substudy comparing HIV-infected adults with slow and rapid disease progress (in the absence of ART) alongside HIV-uninfected controls, increased levels of LPS were found in adults with HIV[99]. However, the slow progressors had lower levels of sCD14 and LBP, and higher levels of EndoCAb (anti-endotoxin core antibodies, which are involved in LPS clearance) than the rapid progressors. This

implies that in those individuals who maintain a low level of viraemia without ART (sometimes termed 'elite controllers'), the immunostimulatory effect of LPS may be diminished[99].

Exogenous administration of LPS to SIV-infected African green monkeys (which have a non-pathogenic infection phenotype), led to an increase in D-dimer levels (indicative of a procoagulant state), sCD14 and a small but significant increase in viral load[101]. LPS may also impact on immune dysfunction by triggering memory B cell death via the Fas/Fas ligand pathway[102].

There have been multiple further investigations into microbial translocation, both mechanistically at the gut barrier and further downstream aiming to detect bacterial products in the blood of HIV infected individuals and any putative impact.

8.4.1 Mechanisms of Microbial Translocation

Mechanistically, there appear to be several pathways leading to impaired intestinal barrier function (Figure 8). In adults with untreated HIV, there appears to be decreased epithelial resistance as assessed by both impedance spectroscopy and [^{3}H]mannitol fluxes compared with HIV infected but treated adults or HIV-uninfected controls[103]. This was associated with increased evidence of apoptosis: upregulation of the pore-forming claudin-2, and mucosal production of interleukin-2 (IL-2) and tumour necrosis factor-α (TNFα)[103](Figure 8). Massive CD4 T cell depletion in the lamina propria is mirrored by gut epithelial cell apoptosis, potentially driven directly by the virus itself[94,104]. The mucosal CD4 T cell population in the gut appears to be preserved if ART is initiated during acute infection, and shows better recovery if ART initiation occurs at a threshold of >350 CD4 T cells/mm^3 compared to lower CD4 counts [105]. There also appears to be a reduction of tight-junction proteins in intestinal epithelial cells among those infected with HIV compared with controls, more marked in the distal than proximal colon, and in those with incomplete CD4 T cell recovery

versus those with better immune restoration[106,107]. *In vitro,* Nazli *et al.* demonstrated a decline in transepithelial resistance in epithelial cells in response to exposure to HIV envelope protein gp120 with associated disruption of tight junction proteins and upregulation of inflammatory cytokines, with possible further disruption to barrier function[95].

Poor response to ART is also associated with decreased epithelial proliferation despite neutrophil infiltration (representing epithelial breaches) in the gut of adults with HIV in the USA[108](Figure 8). T-helper 17 (Th17) cells, a CD4 T-cell subset necessary for mucosal defence also appear to be preferentially depleted in pathogenic SIV infection as compared to non-pathogenic SIV, where Th17 mucosal frequencies are unaffected[109,110]. Th17 cell numbers and function both appear to decline with more advanced HIV infection and concomitant increasing immune activation[111].

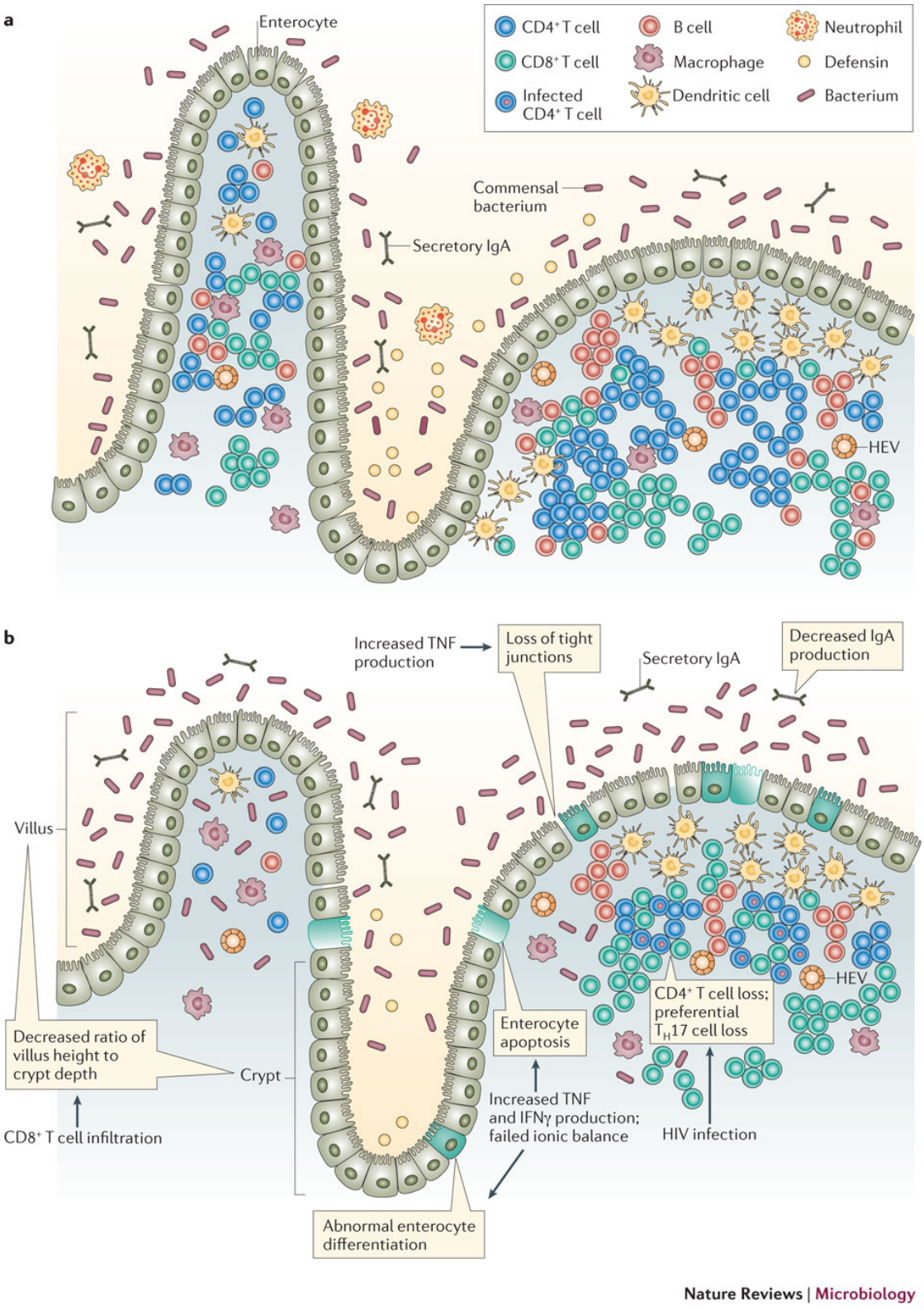


Figure 8 The intestinal epithelium in health and during HIV infection, reproduced with permission from Sandler *et al.*, 2012[112]

The intestinal epithelium in a healthy individual. A continuous enterocyte lining with intact tight junctions prevents translocation of commensal bacteria from the intestinal lumen into the intestinal wall. Neutrophils (recruited by T helper 17 (Th17) cells in the gut-associated lymphoid tissue (GALT)), defensins (produced by Th17 cells) and secretory immunoglobulin A (IgA) maintain control over the growth of commensal bacteria, further impeding microbial translocation[113]

The intestinal epithelium in an HIV-infected individual. The villus height in the intestine of an HIV-infected individual decreases with an increase in crypt depth[114,115] in association with CD8 T cell infiltration[116]. The decreased villus height/crypt depth ratio has been attributed to abnormal enterocyte differentiation[117] and enterocyte apoptosis, which in turn may be caused by failure of the cells to maintain ionic balance and by increased production of interferon-γ (IFNγ) and tumour necrosis factor (TNF).

Increased TNF production may also lead to the destruction of the tight junctions[118]. B cell dysfunction may contribute to decreased luminal IgA concentrations[114,119], and HIV infection of CD4 T cells is likely to drive the loss of CD4 T cells[120-122], particularly Th17 cells, from the GALT. Lower IgA levels and Th17 cell loss may allow bacterial overgrowth, which may also contribute to increased microbial translocation[123]. The continued presence of the high endothelial venules (HEVs) suggests there is not an anatomical abnormality preventing T cells from trafficking into the GALT.

There is evidence of increased density of macrophages in the duodenal mucosa of adults with untreated HIV compared with uninfected controls, with associated increased levels of potentially damaging pro-inflammatory molecules such as IL-1β, CCL5, CXCL9 and CXCL10 and reduced phagocytic activity[124]. Impaired ability of intestinal macrophages to phagocytose microbial products has been demonstrated in pathogenically SIV-infected primates, and in humanised mice infected with HIV (Figure 8)[118,125]. Gut mucosal macrophages are usually anergic in response to stimulation by pathogen associated molecular patterns (PAMPs); although they are effective phagocytes, they do not initiate an inflammatory cascade, enabling them to maintain the gastrointestinal barrier without inducing an inflammatory response to the host's intestinal flora[126,127]. Increased macrophage related pro-inflammatory cytokines could have downstream results of inflammation and tissue injury[124]. *In vitro*, incubation of IL-2, IL-4, TNFα and IL-13 all reduced the transepithelial resistance of rat jejunal mucosa[103].

It appears the gut microbiota itself is also altered by HIV infection. Several studies examining stool, rectal biopsies and rectal sponges have found decreased diversity in the microbiota of adults with HIV (mostly on ART), with an increased proportion of *Enterobacteriaceae* and *Proteobacteria*, with a decrease in *Clostridia* and *Bacteroides*[128-132]. In pathogenic SIV, ART initiation was found to decrease the abundance of *Bacteroidetes* and *Firmicutes* with an increase in *Proteobacteria* (and a concomitant increase in the metabolic activity of the *Proteobacteria* as measured by stool RNA), although this normalised over time[133]. *Lactobacillus* decreased dramatically on initial infection, but recovered after one month of ART. One study found increased diversity in stool samples from untreated adults with HIV in the United States (U.S.), with a microbiotal profile more closely resembling that found in agrarian societies than that usually seen in samples from healthy adults in the U.S[131]. Most other studies found a decrease in diversity, more pronounced with more severe immunosuppression, impacting on both stool and mucosal microbiotal profiles[134,135]. One large study of Ugandan adults (82 HIV infected, 40 HIV-uninfected) also demonstrated a reduced richness in the gut microbiome to be associated with poorer immune status, alongside an increase in *Enterobacteriaceae* and a decrease in

Ruminococcus[136]. This study also found an expansion of Adenovirus and *Anelloviridae* within the virome of adults with advanced HIV, the pathological significance of which is as yet unknown.

Bringing together mucosal dysfunction and dysbiosis of the microbiota, Dillon *et al.* described an increased frequency of activated colonic myeloid dendritic cells (mDCs) in HIV, with an associated increased abundance of *Prevotella* species, which *in vitro* had the capability of inducing inflammatory cytokine production from those mDCs. Increased activity of IDO-1 with resulting Th17 depletion was seen in SIV-infected rhesus macaques was associated with a specific depletion of gut *Lactobacillus*[137], and in adults newly infected with HIV, an increased proportion of *Lactobacillales* was associated with lower markers of microbial translocation, higher CD4 T cell percentage and lower viral loads both prior to and during ART[138]. There was some reduction in IDO-1 activity on supplementation with a *Lactobacillus*-containing probiotic[137]. There also appears to be a positive relationship between *Bacteroides* species and the frequency and ability of invariant natural killer T cells (iNKT) in the GALT to produce IL-4 and IL-10, which in turn was associated with less immune activation[139]. With the exception of the Ugandan study[136], numbers have been small (maximum cohort size 40 patients of whom half were receiving ART[135]) and little work has been done investigating the gut microbiome of children with HIV.

Estes *et al*. investigated evidence of microbial translocation in pathogenic versus non-pathogenic SIV infection in rhesus macaques (RMs) and sooty mangabeys (SMs) respectively[118]. Using qualitative image analysis to examine colonic mucosa, there was evidence of increased staining with a monoclonal antibody to lipopolysaccharide (LPS) core antigen in the lamina propria of RMs with chronic SIV infection compared to early acute infection and uninfected controls. LPS staining was also found in both locally draining and remote lymph nodes in chronically infected RMs, as well as in the liver. Using quantitative image analysis techniques involving staining large sections of colonic mucosa, a positive association was found between the percentage of intestinal mucosa staining positive for LPS and that found in local and remote nodes.

LPS staining was co-localised with staining for the innate pro-inflammatory cytokine interferon-α (IFNα) and Interleukin-18 (IL-18), neither of which were found in close proximity to areas of viral replication. Staining for the tight junction protein claudin-3 revealed evidence of discontinuity and epithelial loss in the gut mucosa of chronically infected RMs, but not those uninfected or those in early stage infection (Figure 8). Furthermore, loss of claudin-3 was significantly associated with the amount of LPS within the lamina propria ($r=0.57$, $p=0.032$), supporting the hypothesis that microbial translocation and intestinal barrier dysfunction is a key feature of SIV pathogenesis. Neither infiltration of LPS nor damage to the intestinal barrier was seen in SMs infected with SIV[118]. Furthermore, early initiation of ART may preserve gut barrier function, GALT tissue structure and Th17 cell numbers and poly-functionality[111,140].

8.4.2 Evidence of Microbial Translocation in the Bloodstream

The downstream evidence for and effects of microbial translocation have proved complex to investigate due to controversy about the best methods with which to measure microbial translocation, assay variability and lack of consistency in results seen in different settings. In adults, Jiang *et al.* performed a cross-sectional study in 173 HIV infected patients, along with a longitudinal cohort of 54 followed up for 48 weeks and 15 HIV uninfected controls[98]. LPS was measured using a limulus amebocyte assay and bacterial DNA was measured using a quantitative broad range polymerase chain reaction for the bacterial ribosomal subunit 16S rDNA (16S rDNA PCR). No evidence of 16S rDNA was found in the HIV negative controls. 16S rDNA was negatively associated with CD4 T cell restoration after adjustment for viral load at week 48 in HIV-infected patients, but not before. Those who were on treatment had lower 16S rDNA levels detected than ART naïve patients.

However other studies have given less clear-cut results. Merlini *et al.* studied groups of patients with either partial or no immune response to ART and found divergent CD4 T cell count recovery despite no change in sCD14 or LPS between the groups[141]. An analysis of adults enrolled in the SMART trial (investigating continuous versus

intermittent ART) measured Intestinal Fatty Acid Binding protein (I-FABP) – a small protein released from intestinal epithelial cells after damage – together with LPS, sCD14, EndoCAb and 16S rDNA to elucidate the relationship between these potential markers of microbial translocation and immune activation[142]. sCD14 was positively associated with IL-6, CRP, amyloid A and D-dimer, and I-FABP was higher in HIV infected individuals than in controls. However, there were no differences in 16S rDNA levels between subjects and healthy volunteers. Only sCD14 was associated with mortality, not LPS or any other marker. The authors raised the possibility that host bioreactivity to microbial translocation might be more important than the actual level of translocated organisms. This was mirrored in a Danish cohort where sCD14 was linked to an increased risk of cardiovascular events, but microbial translocation markers were not[78]. Indeed one study demonstrated an independent association of specific CD14 and Toll-like receptor 4 (TLR4) gene polymorphisms with CD4 T cell recovery, although no genome-wide significant determinant of I-FABP or sCD14 were found in 717 untreated adults in a Swiss cohort[143].

In a large prospective cohort of ART-naïve adults in Uganda, markers of immune activation were raised without any evidence of concomitant microbial translocation[144]. Both sCD14 and CRP were raised, and CRP was associated with mortality but there was no change in LPS over time in the absence of ART[145]. More recently, an Italian cohort of HIV infected adults found that CRP was the only biomarker pre-ART to predict non-AIDS events, with sCD14, LPS and EndoCAb all unrelated to disease progression[146]. A French study of early HIV infection in 27 adults found that immune activation occurred in the apparent absence of microbial translocation (as measured by I-FABP and 16S rDNA PCR) in the first 6 months of infection which suggests that microbial translocation is not the primary driver of immune activation at this stage of infection[147]. There was modest association of LPS with an increased ratio of kynurenine to tryptophan (KT ratio) (rho=0.12, p=0.02) linking monocyte activation to microbial translocation in Ugandan adults initiating ART, and an independent association of a high KT ratio with poor CD4 T cell recovery at 1 year and increased mortality[56]. However, another adult cohort study in Uganda

found that only viral load, not T cell activation, was independently associated with morbidity and mortality[148].

8.4.3 Microbial Translocation in Children with HIV

In children the available evidence is more conflicting. Pillakka Kanthikiel *et al.* studied 85 ART-experienced children in India over 48 weeks who were treated with a protease inhibitor in 1997[149]. Potential markers of microbial translocation including 16S rDNA, sCD14 and LPS were compared with markers of immune activation and viral load. Interestingly they reported an increase in all markers of microbial translocation compared with HIV-uninfected adults with no change over time despite 48 weeks of ART. There was an association between levels of 16S rDNA, CD4 T cell count and immune activation but counter-intuitively no association with LPS or viral load, implying that microbial translocation was persisting despite viral control.

Papasavvas *et al.* compared LPS, sCD14, LPS binding protein (LBP) and T cell activation in 54 HIV infected infants enrolled in the South African CHER trial with 22 exposed, uninfected children[150]. LPS was higher in HIV-infected children compared to controls and LPS, sCD14, LBP, and T cell activation were all higher in those infants in whom ART was deferred as compared to those starting immediate therapy. No LPS was detectable in any child between the ages of 6 months and one year. However there was no association between T cell activation and LPS or between the T cell activation rate of change and the LPS rate of change. In a cross sectional Spanish cohort of 77 children with HIV infection, HIV viraemia was linked to T cell activation, but neither LPS or 16S rDNA levels (assessed with a qPCR) were associated with markers of immune activation[151]. The same group compared markers of immune activation and microbial translocation between 10 children receiving standard ART and 5 receiving ritonavir-boosted protease inhibitor monotherapy and found no difference between the groups.

In contrast, Wallet *et al.*, evaluating 33 children starting ART over 96 weeks found that 10% had an elevated LPS, but with no associated rise in sCD14[152]. Although immune recovery was seen along with a decrease in immune activation, there was no association with LPS levels. A raised LPS level was associated with monocyte activation, but not with T cell activation, CD4 T cell count, thymic output or EndoCAb.

8.4.4 Interventions Targeting Immune Activation and Microbial Translocation

There have been numerous studies investigating therapies targeting immune activation and its potential drivers in animal models and adults with HIV, including statins, pro/prebiotics, chloroquine, hydroxychloroquine, cyclooxygenase type 2 inhibitors, leflunomide, intravenous immunoglobulin, minocycline, bovine colostrum, micronutrient supplementation, raltegravir and maraviroc ART intensification, valganciclovir (targeting CMV coinfection), rifaximin, sevelamer, mesalamine, lisinopril, methotrexate and most recently fecal microbiotal transplantation[153]. Results have been mixed, with no single therapy having sustained clinical and immunological impact across settings[154-156]. Sevelamer for example, a phosphate binding drug used with the aim of binding LPS in the gut, appeared promising in reducing immune activation and even viral replication (albeit slightly) in acutely SIV infected pigtailed macaques, but in ART naïve humans, no impact was found on markers of immune activation or microbial translocation[157,158]. A combination of both probiotics and IL-21 to promote Th-17 cell recovery in the gut was reported to contribute to a reduction morbidity in ART-treated, SIV-infected rhesus macaques, although no significant change in microbial translocation was measured, as measured by *E. coli* antigen detection in gut related tissues[159]. Trials in children are lacking. In one randomised double blind controlled trial of 77 children in Brazil, (55% of whom were receiving ART), probiotics (*Bifidobacterium bifidum* with *Streptococcus thermophiles*, 2.5 $x10^{10}$ CFU (colony forming units)) led to greater CD4 T cell rise in the intervention group compared to controls after 2 months of therapy[160].

8.4.5 Challenges in Quantifying Microbial Translocation

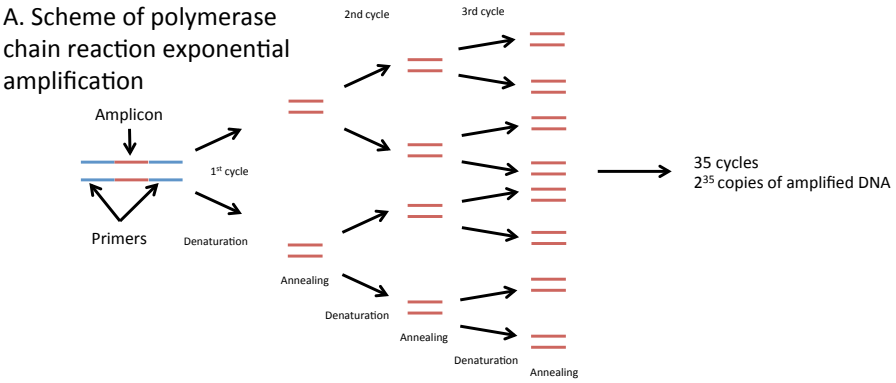
It appears the evidence for microbial translocation causing immune activation in HIV is limited and conflicting particularly in children. This may in part be due to small study sizes, limited prospective data and limited use of appropriate age-matched controls. However, in studies in both adults and children, there is also a lack of consensus about the optimal methodology to test for microbial translocation. LPS is limited in that the ranges obtained in different studies vary by several logs- i.e.

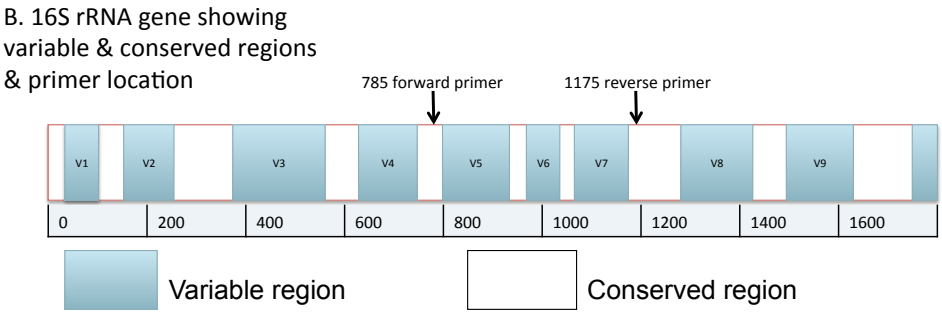
reproducibility is poor[142,144,149]. Plasma inhibition within the Limulus Amoebocyte assay used for quantifying LPS can also be a problem, with different researchers recommending sample dilution from 5 to 100-fold to prevent inhibition, with considerable variation in plasma inhibition between samples within the same study (and even from the same individual)[73,161]. Soluble CD14 is also problematic in that it appears to be influenced by factors other than microbial translocation, including differences in ethnicity[162]. Soluble CD14 is a marker of monocyte activation which can be caused by multiple processes, including raised levels of interferons and by exposure to HIV itself in addition to microbial translocation[163,164]. It has even been shown to be independently associated with smoking in adults with HIV[165,166]. It therefore appears to be a generic marker of immune activation rather than of microbial translocation alone.

8.4.6 Molecular Methods for Investigating Microbial Translocation in Blood

8.4.6.1 Broad Range 16S rDNA PCR

One previously used technique for investigating microbial translocation in blood is a quantitative broad range polymerase chain reaction for the gene of the bacterial 16S ribosomal subunit (16S rDNA PCR) (Figure 9). This involves using primers that target the conserved regions of the 16S ribosomal subunit, aiming to amplify any bacterial DNA present in the sample. The amplicon includes variable regions which can be used to identify the bacterial species present via sequencing techniques. Positive samples are identified via gel electrophoresis (Figure 9) or by a quantitative PCR reaction using a dye such as SYBR Green™ (Life Technologies) to stain double stranded DNA.





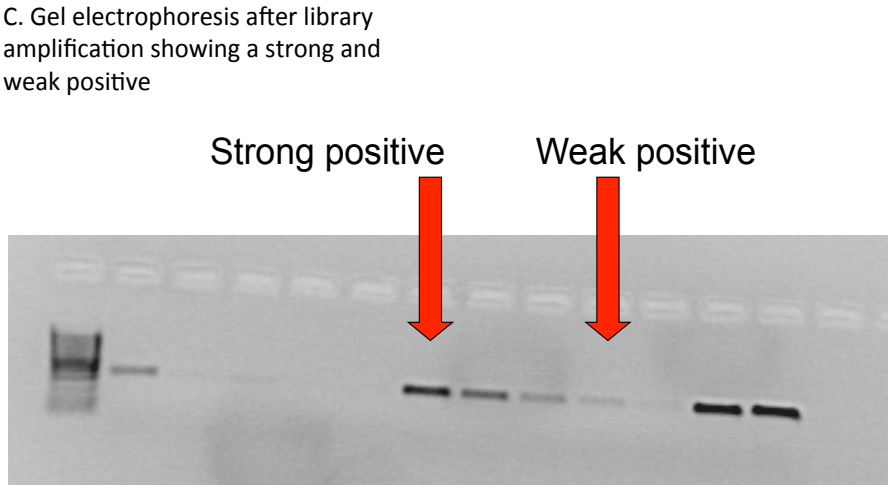


Figure 9 Schema of Broad range 16S rDNA PCR

A: Schematic of principle of polymerase chain reactions, showing sequential denaturation and extension, resulting in exponential amplification of DNA strand

B: Schematic of 16S ribosomal subunit gene showing variable and conserved regions and primer sites

C: Example image of gel electrophoresis post library amplification demonstrating and strong and weak positive result from serial dilutions

It is important to emphasise the methodological difficulties that measuring 16S rDNA presents. 16S rDNA PCR is subject to artefact from endogenous and exogenous bacterial products[167-169] and therefore without sequencing the PCR product, there is concern that changes in the quantity of bacterial DNA detected by PCR may not necessarily be due to gut-derived organisms. In particular, using high numbers of cycles such as reported in studies undertaken by both Jiang *et al.* and Pilakka-Kanthikeel *et al*. will increase the likelihood of false positive results due to contamination[98,149]. Without sequencing, It is possible that differences in results from a quantitative PCR alone might be due to variable copy numbers of the 16S rDNA gene between (and within) species[170].

Additionally, differing levels of host DNA in the blood or plasma can impact on extraction efficiency of bacterial DNA and therefore on bacterial load[171]. Although two studies reported higher levels of 16S rDNA in HIV-infected individuals versus controls[98,149], another found no difference[169], and there has been discrepancy in the associations demonstrated between 16S rDNA and other markers of immune activation and microbial translocation. Several studies have found no relationship between 16S rDNA and other markers such as sCD14, LPS, I-FABP, LPS-binding protein (LBP) or markers of T-cell activation[172-175].

8.4.6.2 Next Generation Sequencing

As described above, most previous studies investigating microbial translocation in blood using molecular techniques have used broad-range 16S rDNA PCR either with or without conventional Sanger sequencing techniques[100,141] (See Figure 10 for Sanger sequencing method). The limitation of this technique is that typically a single dominant species will be identified with minority species often being out-competed in the PCR, or generation of mixed sequences[141] (Figure 10). Cloning can go some way to resolve mixed sequences, but the maximum number of sequences generated would still be in the order of 100, and the process would be extremely labour intensive and expensive. Attempts to identify the 16S rDNA in blood using both

restriction enzyme digestion or Sanger sequencing with cloning has so far yielded results compatible with environmental contamination rather than recognised gut commensals[100,141,176].

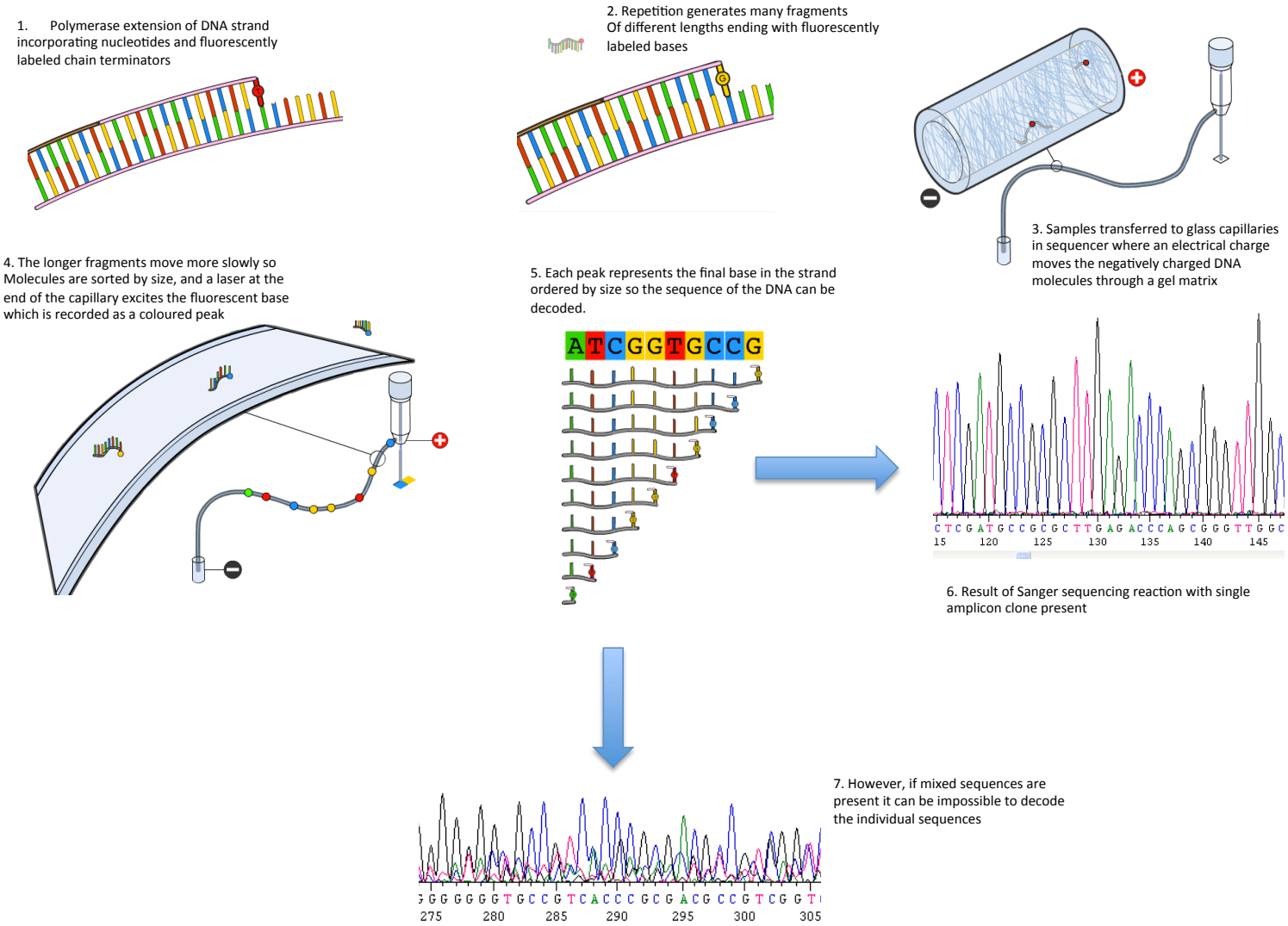


Figure 10 Schematic of Sanger sequencing technique (adapted from Wellcome Trust Sanger Institute animation with permission)

Schematic of Sanger sequencing technique demonstrating stages of strand extension, chain termination, and recording fluorescence. Also showing the difficulty presented by mixed sequencing[177]

High-throughput sequencing (next generation sequencing or NGS) can generate 10^7 sequences from a single sample and give a much more comprehensive description of microbial DNA present in the sample, including those in the minority (Figure 11)[178,179]. This technique involves amplifying a PCR library with barcoded primers, then clean-up and normalization steps followed by the sequencing process on a machine such as a MiSeq™ (Illumina). The massively parallel nature of the process enables rapid and relatively inexpensive sequence generation (as compared with the Sanger method). Utilising paired-end reads allows longer amplicons to be sequenced to the point where operational taxonomic units (OTUs) can be identified to a species level, and dual-indexing the barcodes mean fewer individual barcodes need to be used even when pooling up to 384 samples (Figure 11)[180,181].

After sequencing, downstream bioinformatics processing is necessary to demultiplex samples (allocate barcoded sequences to the samples they derive from), cluster sequences to OTUs, build phylogenetic trees, quality-filter and rarefy the data in preparation for analysis, and perform that analysis. A method which is widely used is the open source software platform Quantitative Insights into Microbial Ecology (QIIME) which 'wraps' various other software dependencies (e.g. the OTU clustering programme uclust and the GreenGenes database which is used as a reference for OTU picking[182])[183] into a single pipeline for analysis[184]. Decisions made at each of these steps can have important implications on the results generated: for example, QIIME supports three methods of OTU picking, *de novo*, closed reference and open reference. *De novo* groups sequences based on sequence identity alone, and can be extremely time consuming. Closed reference compares sequences to a reference database (e.g. GreenGenes) and discards those which do not match, and open reference combines the two other methods, initially comparing with the database, and then *de novo* clustering remaining samples[184]. Therefore these methods are heavily dependent on the quality of the reference database.

To date, one study in pathogenic SIV has used pyrosequencing with 454 FLX technology to sequence 16S rDNA isolated from mesenteric lymph nodes and the

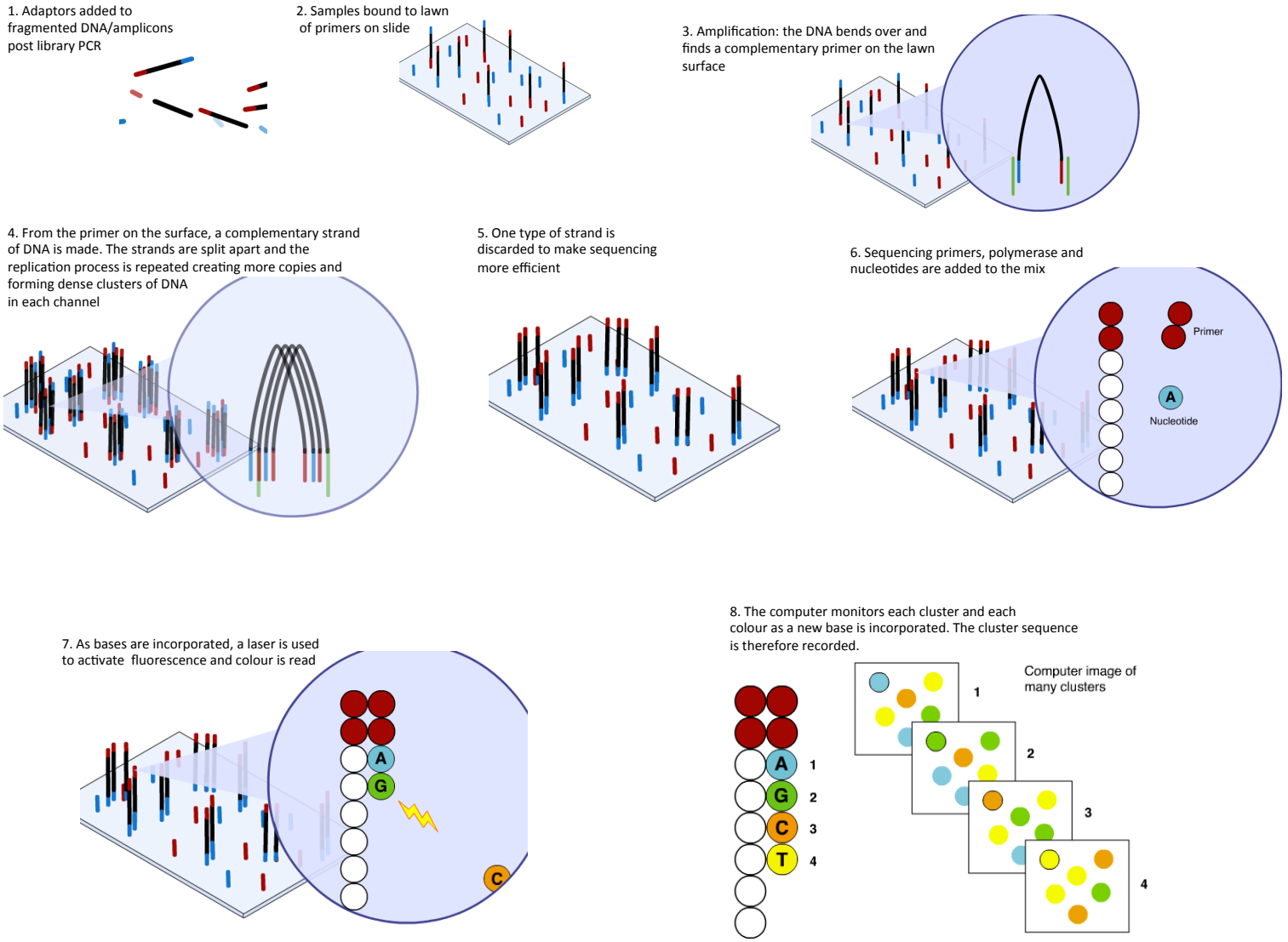


Figure 11 Schematic of Illumina next generation sequencing method (adapted from Wellcome Trust Sanger Institute animation with permission)

Method utilized by MiSeq ™ sequencer showing steps from post-library amplification[185].

Red & Dark Blue: Sequencing primers
Green: barcodes
Pale blue & Orange: 16S primers
Grey: amplicon

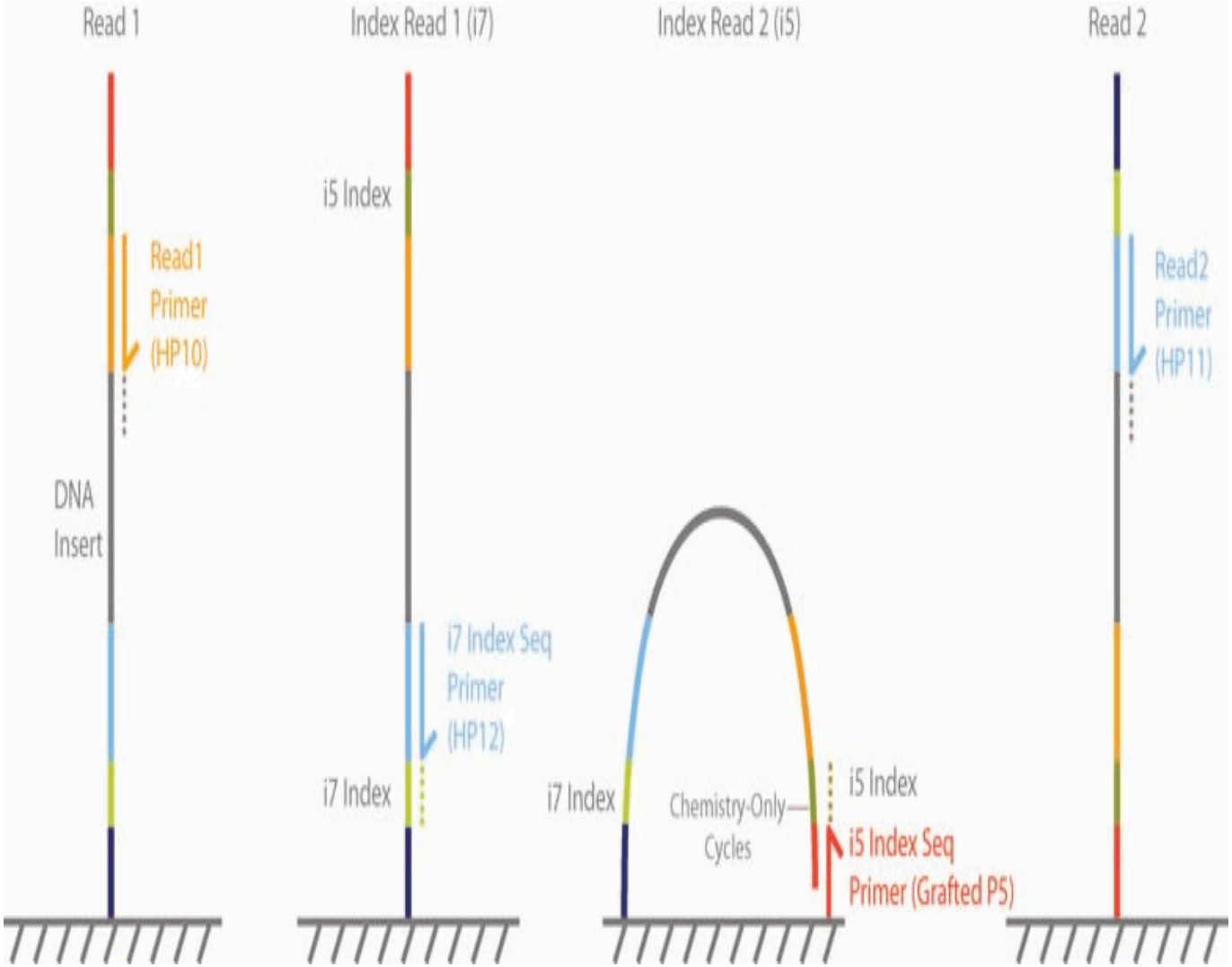


Figure 12 Schema of positioning of dual index bar codes and sequencing primers for Illumina MiSeq™ sequencing method (reproduced with permission from Illumina Indexed sequencing overview guide[181].

liver[133]. This study identified Proteobacteria as preferentially translocating from the gut as evidenced by higher proportions in the tissues studied[133].

The vast majority of studies utilizing next generation sequencing to investigate the human microbiome has focused on bacterial rich samples such as stool, other gut samples, the female genital tract, the nasal cavity or the skin[186,187]. By contrast, studies using NGS to investigate microbial DNA in blood samples are much scantier, and the bacterial load is likely to be several orders of magnitude lower than in stool for example[188-191]. We anticipated that a method designed for studying high bacterial load samples was likely to need modification for lower load samples such as in this study.

8.4.6.3 Real Time/Quantitative Polymerase Chain Reactions

Specific or quantitative polymerase chain reactions (qPCRs) are a well-established method of detecting bacterial DNA (See Figure 13 for schematic of qPCR methodology). They have particular advantages over standard PCRs in that the product (in the form of fluorescence) is detected in real time during the assay, reducing the time to result and quantifying the product. The point at which a qPCR is deemed positive is known as the cycle threshold (CT) and can be used to quantify the amount of DNA present in a sample when compared with controls (Figure 14).

They are also less vulnerable to contamination than broad range PCRs. As broad range 16S rDNA PCRs will amplify any bacterial DNA present in the sample or reagent, they risk being overwhelmed by contamination at a high number of cycles, limiting the sensitivity[167]. qPCRs target specific organisms that are not associated with environmental contamination and so can be made more sensitive by increasing the number of cycles. However, as they are specific, there is the possibility that unpredicted bacteria will be missed. A combination of both a panel of specific and broad range PCRs can provide both sensitivity and specificity in the detection of bacteria that are difficult to grow in culture, such as much of the gut flora[192].

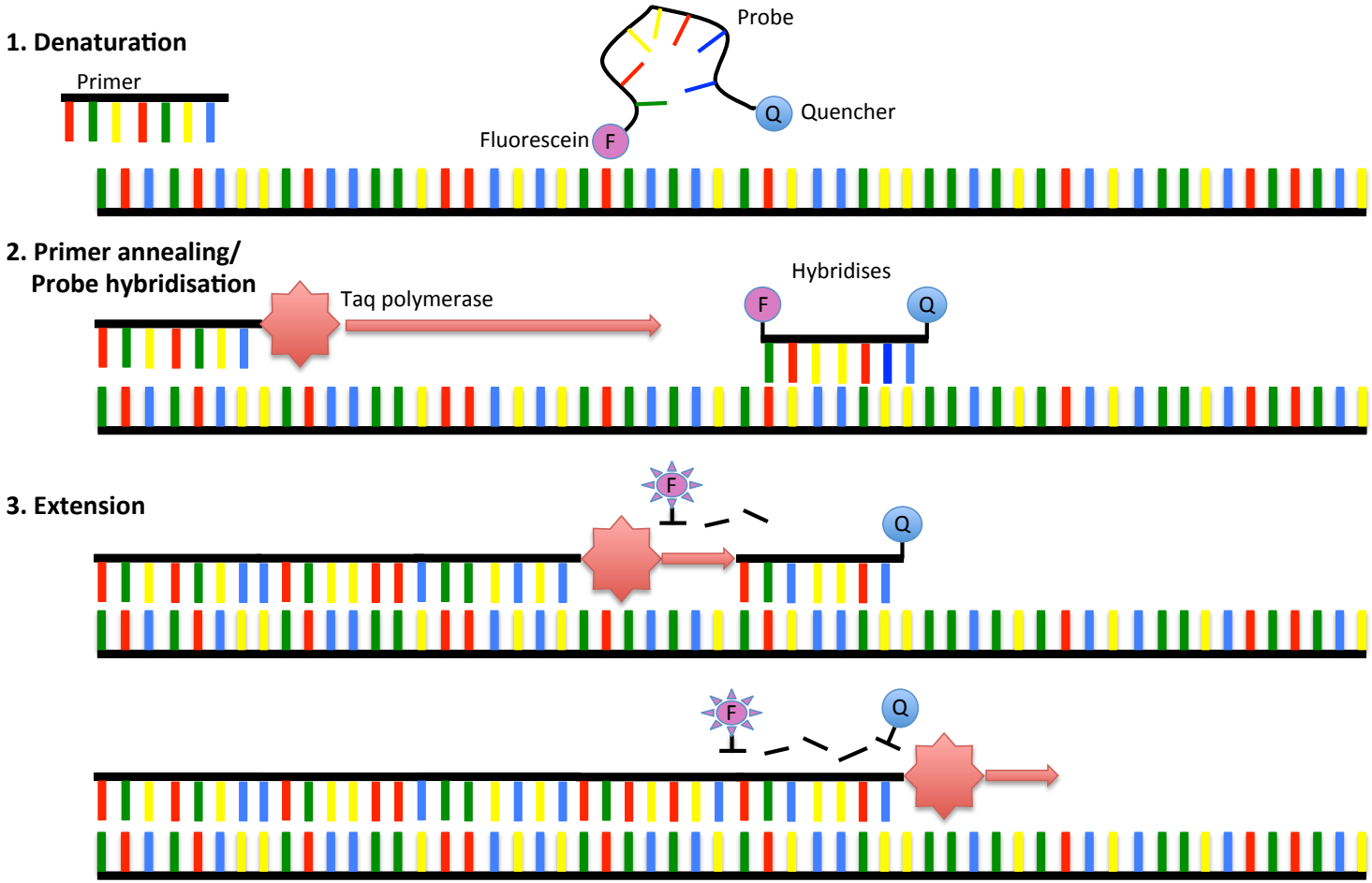


Figure 13 Schematic of qPCR methodology illustrating Taqman™ primer probe method

Cycle Threshold

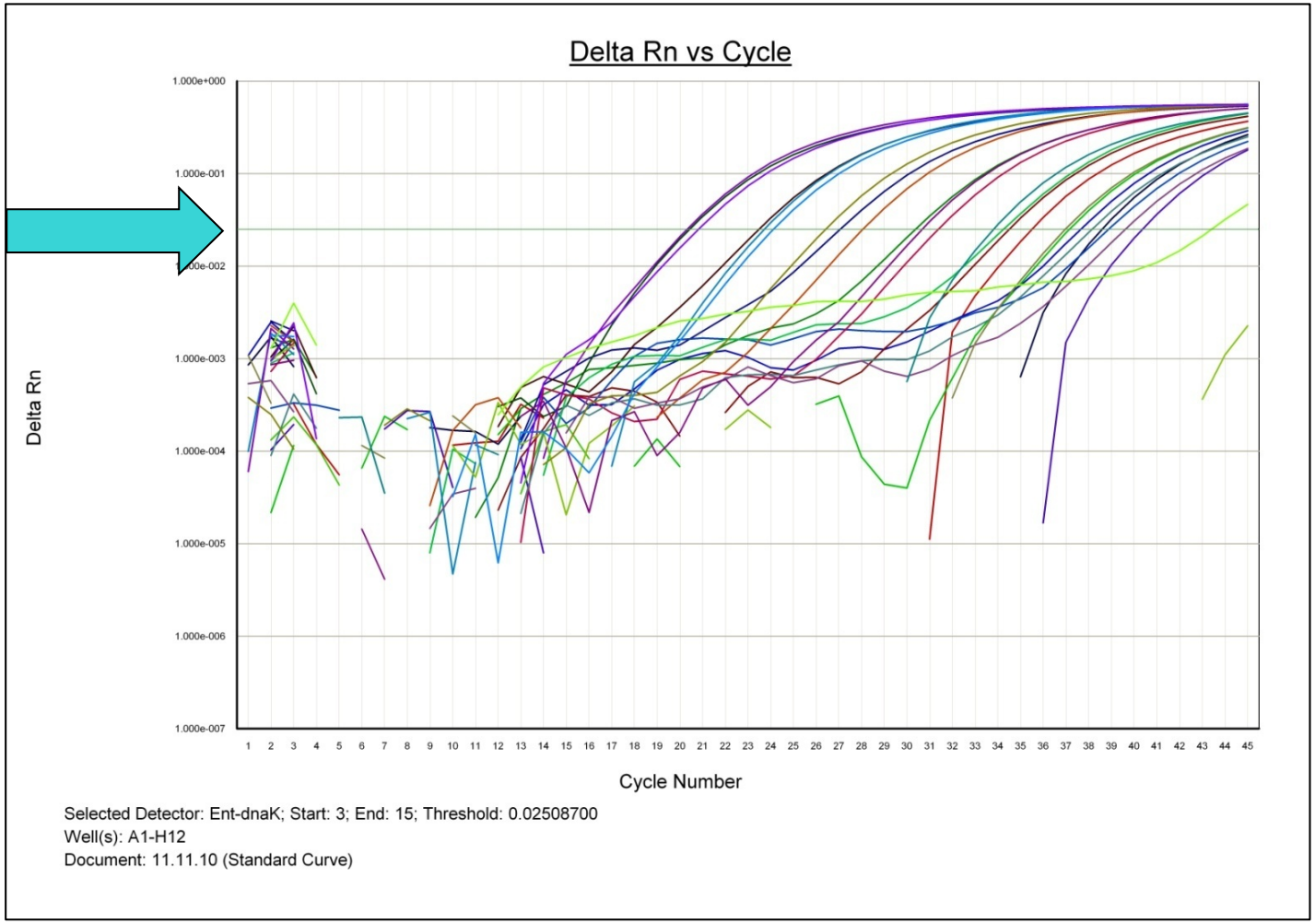


Figure 14 qPCR results showing Cycle Threshold.

Serial dilutions of qPCR assay demonstrating position of cycle threshold for samples with increasing dilution (higher cycle threshold indicates higher dilution and less initial DNA load)

8.4.7 Methods for Assessment of Gastrointestinal Tract Injury

I-FABP is a plasma marker of small intestinal damage used in diverse settings including adult HIV[142,193-195]. In a cohort of 27 adults with early HIV in France, I-FABP was similar at baseline to healthy controls, but increased by month 6 (both in ART naïve and ART treated patients)[147]. In treated adult HIV, raised I-FABP was found to be predictive of mortality along with markers of innate immune activation such as sCD14 and D-dimers[194]. In untreated adult HIV, a strong association exists between I-FABP and sCD14 although not with HIV viral load[196].

HIV-infected children in Africa might be more vulnerable to gut damage than either adults or children with HIV in developed settings, in part due to the contribution of environmental enteropathy. Environmental enteropathy is characterized by intestinal inflammation, villous atrophy (with associated malabsorption) and impaired intestinal barrier function, potentially leading to microbial translocation and downstream effects such as anaemia and stunting that is prevalent amongst those living in unhygienic conditions[197]. This could manifest in more pronounced stunting and wasting in HIV-infected children, even than uninfected peers from the same community. Investigating I-FABP, microbial translocation markers and anthropometry could help to tease out the relative contribution of HIV and microbial translocation to gut damage, stunting and wasting.

8.5 Pilot Study

I undertook a cross-sectional pilot study using broad range 16S rDNA PCR with standard Sanger sequencing (See section 8.4.6.1) and developed new assays to identify microbial 16S rDNA in the blood of HIV-infected children attending Great Ormond Street Hospital (GOSH), London.(Appendix E Paper: “Evidence Microbial Translocation Occurs in HIV-Infected Children in the United Kingdom”; AIDS Research and Human Retroviruses; October 2013) Discard whole blood EDTA samples were collected from children attending the HIV outpatient clinic at GOSH. Broad range 16S rDNA PCR was carried out using two primer sets[167]. 16SFa/16SFb

(GCTCAGATTGAACGCTGG/ GCTCAGGAYGAACGCTGG) and 16SR (TACTGCTGCCTCCCGTA) amplified the V1 and V2 region of the 16S rDNA gene and 785F (GGATTAGATACCCBRGTAGTC) and 1175R (ACGTCRTCCCCDCCTTCCTC) amplified the V5 and V6 region. Amplicons derived from positive samples were sequenced using the Big-Dye v.3.1 cycle sequencing kit (Applied Biosystems, Warrington, UK) and analyzed on the 3130 genetic analyzer (Applied Biosystems). A subset of samples found to be positive by broad range 16S rDNA PCR was analyzed by high-throughput sequencing. Attached to the 16S rDNA primers were standard 454 Titanium adapters and an individual barcode sequence for each sample. The library was pyrosequenced on a 454 FLX Titanium (Roche) platform according to the manufacturer's recommended protocol. Sequences were processed and analyzed using QIIME[198]. 16SFa/16SFb/16SR and 785F/1175R reads were discarded if they contained ambiguous bases, if the quality score was < 25, if the run of homopolymer bases was > 6, if there was a mismatch in primer sequence, and if the barcode could not be corrected. Operational taxonomic units (OTUs) were clustered at 97% sequence similarity and chimeric sequences were removed using USEARCH[182]. Representative OTUs were assigned taxonomy using the RDP classifier[199] at a minimum support threshold of 80% again the Greengenes database[183]

Samples were obtained from 105 sequential children attending the outpatient clinic, representing 85% of the clinic population. 9 samples were positive using broad range 16S rDNA PCR, a positivity rate of 8.6%. Of the nine patients in whom microbial DNA was detected, five (55%) had an undetectable viral load; four (45%) had a CD4 percentage above 30% and two (22%) had a CD4 of less than 10%. In immunocompetent children of equivalent age, the positivity rate was less than 1%. From 3 of these amplicons, 16S rDNA sequences were identified using Sanger sequencing as *Streptococcus* species (mitis group); *Propionibacterium acnes* and Coagulase negative *Staphylococcus*. Direct Sanger sequencing failed in 6 samples, probably due to a mixture of 16S rDNA sequences from different species. These results highlight a major disadvantage of standard broad-range 16S rDNA-PCR in that only the dominant sequence is identified. NGS of 4 samples identified the same species as obtained by Sanger sequencing and in addition identified 10 more species

of differing proportions in each of the 4 samples, including known gut commensals (*Bifidobacteriaceae, Lactobacillaceae), Staphylococcaceae* and *Streptococcaceae*. This was the first published application of deep sequencing to investigate microbial translocation in HIV[200].

It would thus appear that microbial translocation may occur in approximately 1 in 10 HIV infected children in London. We hypothesized rates could be higher in children in Africa, where environmental enteropathy is likely to be more prevalent.

8.6 Summary

Although ART has revolutionised the management of the HIV pandemic, we face new challenges in supporting children with perinatal HIV infection through a lifetime of therapy. To minimise the risk of comorbidities in the future for these children, efforts must be made to investigate potential causes of those morbidities such as immune activation. It is hypothesised that HIV damages GALT, allowing translocation of gut microbiota and microbial-associated products, with increased circulating levels of LPS and bacterial ribosomal DNA. This could drive immune activation and T-cell dysregulation, and may increase morbidity and mortality in patients with and without ART. This effect is thought to be so important that clinical trials are underway to reduce microbial translocation aiming to reduce immune activation and improve clinical outcome[155,157,201-203].

In summary analysing the results of work undertaken to date, a number of issues arise:

1. Most studies have been cross-sectional, not prospective[100,204-206]. Translocation may be due to severe HIV disease/immunological non-response, and not causally implicated[207].

2. Many studies utilise quantitative PCR (qPCR) to quantify the 16S rDNA subunit to measure translocating bacteria. 16S rDNA PCR is subject to artefact from endogenous and exogenous bacterial products[167-169] and therefore without sequencing the PCR product, there is concern that results may not be due to gut-derived organisms. Sequencing the 16S rDNA has so far yielded results compatible with environmental contamination rather than recognised gut commensals[100,141,176].

3. Some studies have failed to find evidence of microbial translocation at all, or report conflicting results depending on markers used[144,145,202].

4. There are few data from Africa: one major African longitudinal study found no significant increase in markers of microbial translocation or circulating cytokines during untreated disease progression in adults[144].

5. Data in children are limited and conflicting: two small studies showed an association between LPS, 16S rDNA and virological response, while another failed to demonstrate a relationship[73,152,208].

Whilst microbial translocation is a plausible hypothesis, and potentially amenable to therapeutic interventions, there are potential pitfalls with technology used to date and inconsistent data. Furthermore, few studies have investigated microbial translocation in HIV-infected children, and none in Africa. It is unclear if and how microbial translocation affects outcome in HIV-infected children.

8.7 Aims of the Thesis

The key aims of the thesis were:

1. To assess the level of microbial translocation defined as identification of bacterial DNA, occurring in HIV-infected African children by both HIV-infected

ART-naïve and ART-experienced, compared with HIV-uninfected African controls

2. To determine the effect of ART on microbial translocation by comparing levels detected in ART-naïve and ART-experienced children, and changing levels over time on ART.
3. To determine the impact of environment on microbial translocation by comparing levels in ART-naïve HIV infected children in urban versus rural settings at baseline and over time on ART.
4. To investigate the relationship between microbial translocation, gut damage, and clinical events in HIV infected children in Uganda.

8.7.1 Primary Hypothesis

1. HIV infected ART naïve children have higher levels of microbial translocation than age matched HIV-uninfected controls.

8.7.2 Secondary Hypothesis

1. ART can reduce microbial translocation in HIV-infected children during the first 72 weeks of therapy.
2. Environment (rural versus urban) has a significant impact on microbial translocation in HIV-infected children.
3. There is a significant relationship between microbial translocation, gut damage, nutritional status, and clinical events in HIV-infected children on ART.

9 Chapter 9 Materials and Methods

9.1 Study Population

The study population comprised 196 HIV-infected Ugandan children (median age 3 years; range 2 months to 11 years) enrolled in the CHAPAS-3 trial, a three-arm phase II/III open-label randomised trial comparing toxicity and pharmacokinetics of three fixed-dose combination ART regimens in children in Uganda and Zambia (ISRCTN69078957). The trial enrolled 480 ART-naïve and ART-experienced children and randomized them to either stavudine (d4T), zidovudine (AZT) or abacavir (ABC) with lamivudine (3TC) and either efavirenz (EFV) or nevirapine (NVP) (Figure 15). The trial results were published in February 2016, and showed good clinical, virological and immunological responses in all three arms with low toxicity[209].

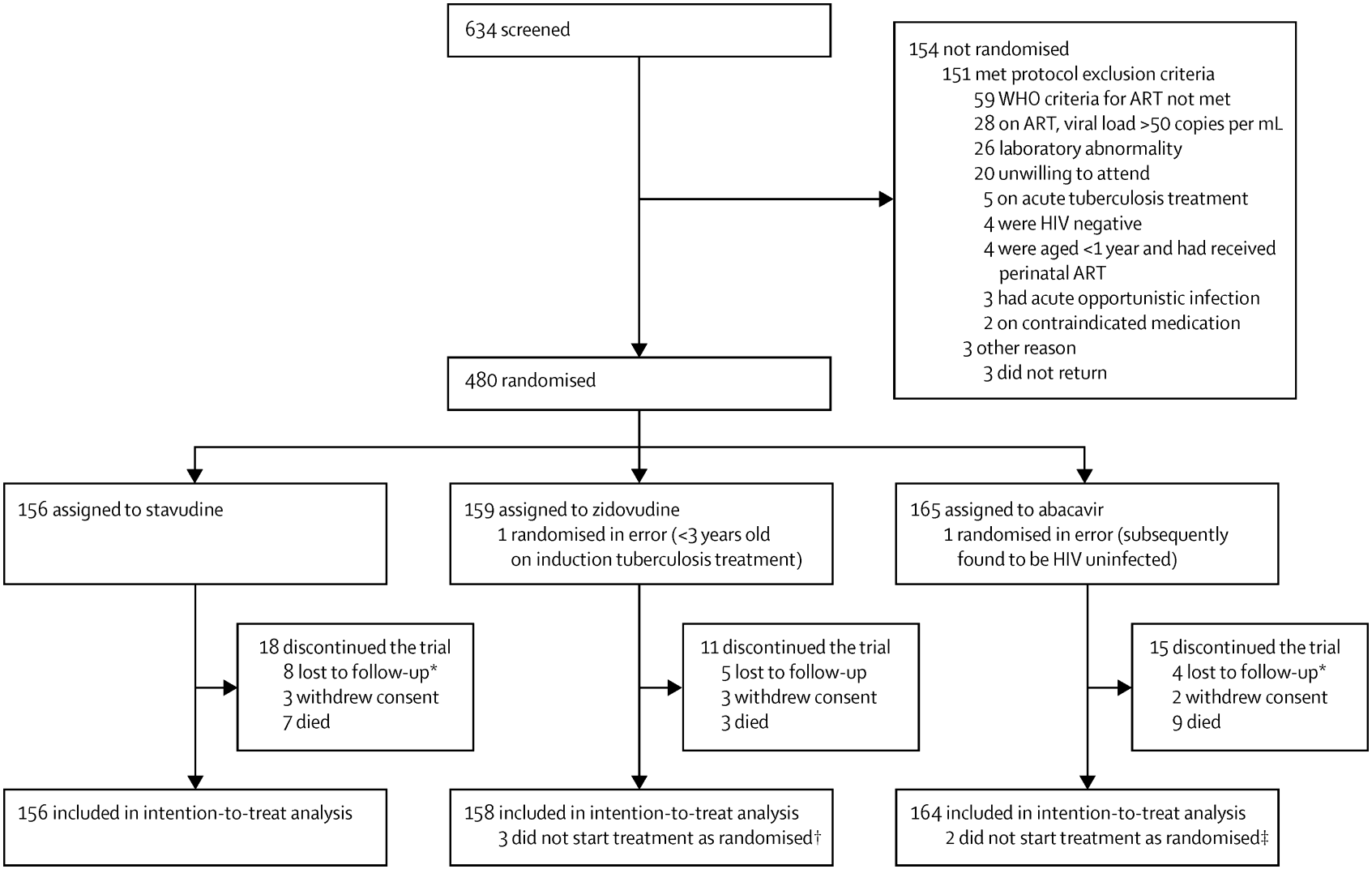


Figure 15 Trial Profile from CHAPAS 3 reproduced under Creative Commons License from Mulenga *et al.* 2016[209]

ART: antiretroviral treatment. *Includes one not seen after randomisation. †One participant started stavudine and substituted zidovudine at 12 weeks, two started abacavir and did not change (both prescribing errors). ‡Two started zidovudine and did not change (one prescribing error and one child changed regimen to match twin sibling).

To be eligible for enrolment, ART experienced children had to have been stable on d4T-containing first line ART for at least two years with a viral load at screening <50 copies/ml, stable CD4 T cell counts/% and no signs of lipodystrophy. All children initiated cotrimoxazole if they were not already on it (dapsone was used as an alternative if necessary). Age, ART, laboratory tests, anthropometric measurements (weight, height and BMI-for-age Z-scores) and co-morbidities were recorded longitudinally until 01/14. Weight-for-age, height-for-age and BMI-for-age z-scores were calculated using UK 1990 cohort data as WHO z-scores did not cover the age range of CHAPAS-3 children for weight[210].

Children were reviewed every 6 weeks by a nurse and every 12 weeks by a doctor for clinical review, and checking self-reported adherence. Haematology, biochemistry and CD4 T cell tests were performed at weeks 6, 12, 24 and then 24-weekly (results available to clinicians), and plasma saved for viral load which was performed retrospectively (viral load measurements were not part of standard care at the time; results not available to clinicians in real time). ART substitutions for toxicity or switching to second line for ART failure was at the managing physician's discretion. The primary outcome was grade 2 or greater clinical adverse events, confirmed grade 3 laboratory adverse events or any grade 4 laboratory events. Adverse events were reported by local trial teams to MRC CTU and clinical adverse events were adjudicated by an endpoint review committee blinded to randomisation. Viral load was tested on stored samples with the Roche COBAS Ampliprep/Taqman version 2.0 and run with a 1/5 dilution of Basematrix 53 giving a lower limit of detection of 100 copies/ml. If children needed to be admitted to hospital, an additional form was filled out by a nurse to document the events of the admission as available from the local hospital. Any additional microbiological samples (e.g. blood cultures or ear swabs) taken by the doctors or nurses in the trial clinic were also recorded.

In terms of the samples used for this study, samples were collected from 141 HIV-infected Ugandan children from the Joint Clinical Research Centre (JCRC) Kampala (urban) and 55 from Gulu (rural): at study enrolment 119/55 were ART-naïve and

22/0 were ART-experienced with suppressed HIV viral load. Children had frozen plasma and pellets stored at baseline then 12 weekly. Pellets and plasma were separated by centrifugation within two hours of collection at 1500g for 15 minutes, and then stored at -80°C. Any discard of blood taken for diagnosis of acute illness was stored. Also included were a cross-sectional sample of 109 HIV-uninfected children from community clinics in Kampala, age-matched to the HIV-infected Kampalan children, who were enrolled as a parallel cohort in CHAPAS-3, assaying a single time point using 0.5 ml of 10 mL frozen plasma.

9.1.1 Sample Size Calculations

The sample size for Aim 1 (primary objective) was 119 ART-naïve HIV-infected children and 90 HIV-uninfected age matched controls and 22 ART-experienced children with 20 age matched HIV-uninfected controls, all from Kampala. This provided >80% power to detect 15-20% differences in prevalence of microbial translocation between ART-naïve HIV-infected and uninfected children based on previous estimates[211]; power was lower for children already on ART. For aim 2, plasma already stored from three time points was assessed for HIV-infected children: weeks -2, 12, and 72 after enrollment in CHAPAS 3/initiation of ART. Aim 3 compared 55 HIV-infected ART-naïve children from Gulu with 119 equivalent urban children from Kampala (JCRC site), which provided ≥80% power to detect differences of 20-25%, determining if results from Kampala were broadly representative of other populations or if there were major effects of environment/diet.

9.2 Ethical Considerations

The study was approved by both UCL Research Ethics Committee (reference 5019/001) and the Ugandan National Council for Science and Technology (reference HS 1559). Informed consent was obtained for the storage and use of samples for studies investigating the pathogenesis of HIV.

9.3 DNA Extraction for Molecular Testing

Pellets were spun from 1 ml of whole blood as per section 9.1. The whole sample available was used. 200 µL of lysis buffer was added and extracted using a Qiagen QIAamp DNA Mini Kit with additional bead beating step. The final elution volume was 200 µL. Plasma samples (200 µL) were extracted using EZ1 robot and EZ1 Virus Mini Kit v2.0 again with additional bead beating step and eluted to a volume of 150 µL. All samples were extracted on first thaw, had an internal positive control (IPC, mouse genomic DNA, non-coding region) added prior to the bead-beating step. Negative extractions were included with each batch as a negative control.

9.4 Quantitative PCR

The steps undertaken to design, test and use novel qPCR primers were as follows:

i. Identification of target organisms
ii. Identification of target genes for amplification
iii. Designing primer and probe sets specific for the target gene, incorporating variability within genus/class while excluding other non-target organisms
iv. Testing the primer probe mix on samples with known bacterial strains
v. Testing experimental samples using the primer probe mixes

9.4.1 Identification of Target Organisms

Target organisms were identified using previously published human microbiome data. Organisms were selected in descending order of priority:

i. Previously identified in blood in studies of microbial translocation in people with HIV
ii. Previously identified in other body tissues in studies of microbial translocation in people with HIV

iii. Previously identified in stool in studies of microbial translocation in people with HIV OR identified in blood/other body tissues in other inflammatory diseases where microbial translocation is implicated in pathogenesis

iv. Previously identified in studies of malnutrition in children in African settings

v. Previously identified as important components of the healthy human microbiome

9.4.2 Identification of Target Genes for Amplification

The primary strategy for identifying target genes was to identify a gene that is specific only for the intended organisms and find well conserved ubiquitous regions within that gene as this minimises overlap with other species. A secondary strategy if the first failed was to target a ubiquitous gene (such as the 16S rDNA gene) and identify an area within the gene that is specific for the intended organism.

9.4.3 Designing Primer Probe Sets

Once target genes were identified, multiple sequences from different bacterial species within the target genus were aligned using MegAlign (Lasergene10, DNAStar). Conserved areas were then entered into Primer Express (Thermofisher), altering parameters to allow a minimum amplicon length of 100 base pairs (to allow sequencing). If no suitable matches were found using the most stringent settings, parameters were relaxed to allow different melting and annealing temperatures.

9.4.4 Testing Primer Probe Sets on Samples with Known Bacterial Strains

The following primer probe sets were used:

Table 1 *Bifidobacterium* (*xfp*) primer probe set

Forward (244-263)	5' CAGCTBTCCGAGCACCAGAT
Reverse (325-343)	5' TCCTTCGTWCACGTGATCG
Probe (270-285)	5' FAM- CTTCCTCGAGGWCTAC-MGB

Table 2 *Lactobacillus* (*hsp60/GroEL* gene) primer probe set

Forward: 223-238 16bp	5' GCTGASGCHATGGAAA
Probe (casei): 242-256 15bp	5' FAM-TTGGSCASGATGGTG-MGB
Probe (acidophilus): 242-256 15bp	5' FAM-TTGGTMACGATGGTG-MGB
Reverse (casei): 301-322 22bp	5' GTAGTTGAAGGTATGCAATTSG
Reverse (acidophilus): 301-322 22bp	5' GTTGTTGAAGGDATGCAATTSG

Table 3 *Clostridiales* primer probe set

Target	*Primer*	*Sequence*
Clostridium leptum subgroup[212]	sg-Clept-F	5' GCACAAGCAGTGGAGT
	sg-Clept-R3	5' CTTCCTCCGTTTTGTCAA
Clostridium coccoides group[213]	g-Ccoc-F	5' AAATGACGGTACCTGACTAA
	g-Ccoc-R	5' CTTTGAGTTTCATTCTTGCGAA

Reaction mix 1 (*Bifidobacterium*): 10µL Taqman® Real Time PCR Master Mix (Life Technologies), 1µL of each of forward and reverse primers and probe (10 pM) and 7µL sample using a 7500 Fast Real Time PCR system (Life Technologies). Cycling conditions were 95°C for 5 minutes then 45 cycles of 95°C for 30 seconds and 60°C for 30 seconds.

Reaction mix 2 (*Lactobacillus*): 10µL Taqman® Real Time PCR Master Mix (Life Technologies), 1µL forward primer and 0.5µL each of the reverse primers and probes (10 pM) and 7 µL sample using a 7500 Fast Real Time PCR system (Life Technologies). Cycling conditions were 95°C for 5 minutes then 45 cycles of 95°C for 30 seconds and 60°C for 30 seconds.

Reaction mix 3 (*Clostridiales*): 12.5µL Power SYBR® Green PCR Master Mix, 0.5µL forward and reverse primers (20µM) and 10.5µL PCR grade water (Bioline) to make a

total volume of 25µL. Cycling conditions were 95°C for 10 min 1 cycle, then 95°C for 15s & 50°C for 1 min (repeated 40 times), then final dissociation stage 95°C 15s 50°C 1 min 95°C 15s.

9.4.5 Testing Experimental Samples using the Primer Probe Mixes

Reaction mixes for individual assays were made up as below, with 2uL of molecular grade water (Bioline), 10µL of Quantifast Multiplex PCR+R master mix (Qiagen) and 5µL of sample. All assays were performed on an Applied Biosystems 7500 Real-Time PCR System™ (Thermofisher) according to the Fast protocol. Pellets were found to have significantly higher cycle thresholds for the internal positive controls than the plasma samples (Two-tailed p value<0.0001) which indicates template inhibition. Therefore they were run at a 10-fold dilution.

Table 4 Reaction constituents for qPCR reactions

Target	Reagent	Sequence	Sensitivity (CFU/ reaction)	Concentration/ Volume
Mix 1				
Staphylococcus aureus [214]	*coa*-F	5'-GTAGATTGGGCAATTACATTTTGGAGG	0.1-1	0.15uM
	coa-R	5'-CGCATCTGCTTTGTTATCCCATGTA		0.15uM
	coa-Probe	5'FAM- TAGGCGCATTAGCAGTTGCATC-BHQ1		0.15uM
Streptococcus pyogenes[215]	*csrR*-F	5'-TGGATGTGGTTGCAGGTTTAGAC	0.1-1	0.3uM
	csrR-R	5'- CGGGCAAGTAGTTCTTCAATGG		0.3uM
	csrR-Pr	5'-JOE-CGGTGCAGACGACTATATTGTTAAACC-BHQ1		0.2uM
IPC[215]	*Mus*	5'-GGACACTATGCCCCTCCTTAGA		0.1uM
	Mus	5'-AGCTCCAAACTCCGTCTCTGTAA		0.1uM

	Mus	5'Cy5-TTGGGAACAAAACACCCATGGAAGGA-BHQ2		0.1uM
Mix 2				
Enterobacteriacae[215]	F *dnaK*	5' ACCTGGGTACWACCAACTCTTGTGT	1-10	0.25uM
	R *dnaK*	5' GTCACTGCCTGACGTTTAGC		0.25uM
	Probe *dnaK*	5'-JOE-AGGATGGTGAAACTCTGGTWGGTCAGCC-BHQ1*		0.25uM
Staphylococcus species [215]	F *tuf*	5' CATTCCAACTCCAGAACGTGAYT	1-10	0.1uM
	R *tuf*	5'-CACGACCAGTGATTGAGAATACG		0.1uM
	Probe *tuf*	5'-CY5-TGAYAAACCATTCATGATGCCAGTTGAGG-BHQ2		0.1uM
Mix 3				
Fusobacterium spp.[216]	F	5' GGATTTATTGGGCGTAAAGC	0.5-5	0.1uM
	R	5'GGCATTCCTACAAATATCTACGAA		0.1uM
	Probe	JOE 5'CTCTACACTTGTAGTTCCG BHQ1		0.1uM
Mix 4				
Bifidobacterium spp.	*xfp*-F	5' CAGCTBTCCGAGCACCAGAT	0.5-5	0.5uM
	xfp-R	5' TCCTTCGTWCACGTGATCG		0.5uM
	xfp-Pr	5' FAM- CTTCCTCGAGGWCTAC-MGB*		0.5uM
Mix 5				
Lactobacillus spp.	F	5' GCTGASGCHATGGAAA	1.3-13	0.5uM
	R (casei)	5' GTAGTTGAAGGTATGCAATTSG		0.25uM
	R (acidophilus)	5' GTTGTTGAAGGDATGCAATTSG		0.25uM
	Probe (casei)	5' FAM-TTGGSCASGATGGTG-MGB		0.25uM
	Probe (acidophilus)	5' FAM-TTGGTMACGATGGTG-MGB		0.25uM

Table 5 Cycling conditions for qPCR reactions

95°C	5 minutes	1
95°C	30 seconds	45
60°C	30 seconds	

9.5 Broad Range 16S rDNA PCR

The reaction mix consisted of 5µL sample, 10.5µL PCR grade water (Bioline), 12.5µL Power SYBR Green master mix (Life Technologies) and 0.5µL 785F (5' GGATTAGATACCCBRGTAGTC, 20pM) and 1175R (5' ACGTCRTCCCCDCCTTCCTC, 20pM) amplifying the V5 – V7 region of the 16S rDNA gene[200]. Each run included multiple dilutions of a standard of known concentration and negative controls in triplicate. Pellets were run at a 10-fold dilution as per section 9.4.5. Cycling conditions were as follows:

1. 95°C for 10 mins
2. 95°C for 15s, 60°C for 1 min (40 cycles)
3. Dissociation: 95°C for 15s, 60°C for 1 min, 95°C for 15s

By extrapolation from standards of a known concentration (*Escherichia coli*), bacterial load defined as colony forming units per PCR reaction (5µL) were calculated (CFU/PCR reaction). CT values generated by negative extraction samples or negative controls in each run were taken as a cut off of zero, meaning that any samples with a higher CT value were considered negative as the assay had reached its limit of detection. By this method, the sensitivity of the assay was calculated as a bacterial load of 5-50 CFUs/PCR reaction.

9.6 Sanger Sequencing

Amplicons were sequenced using the Big-Dye v.3.1 cycle sequencing kit (Thermofisher) and analyzed on the 3130 genetic analyzer (Thermofisher). The sequences obtained were compared to those on the GenBank database using the BLAST program available at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). The sequence was classified to species level if there was > 98% homology with two or more GenBank sequences from the same species, submitted by independent laboratories, and the percentage identification was lower for all other species.

9.7 Next Generation Sequencing

Library preparation used reagents as per Table 6 using the same primers with individual 12-bp error correcting barcodes/indices (different for forward and reverse to allow paired-end sequencing), and Illumina compatible adaptor, pad and linker sequences (See Appendix B Tables of barcoded primer sequences for NGS)[198,217]. Negative extractions and negative controls were included for quality control as well as mock communities (obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community A (Even, Low Concentration), v3.1, HM-278D, and Microbial Mock Community B (Staggered, Low Concentration), v5.2LHM-783D (Appendix A Species in Microbial Mock Communities)

Primer sequences were checked against 16S rRNA gene sequences of species within microbial mock communities using Megalign™ to ensure no primer bias against these species.

Cycling conditions were as per Table 7.

Table 6 Reagents for NGS library amplification

Reaction component	Final Concentration	Per 25 uL reaction
PCR grade water	-	15.2 -16.2 uL
Molzym PCR buffer	1X	2.5 uL
dNTPs	200pmol/ul (25pmol/ul working concentration)	0.2 uL
Forward primer (with barcode attached, different for every sample)	0.4pmol/ul (10pmol/ul working concentration)	0.5-1 uL
Reverse primer (with barcode attached, different for every sample)	0.4 pmol/ul (10pmol/ul working concentration)	0.5-1 uL
Moltaq	0.025pmol/ul (5pmol/ul working concentration)	0.125 uL
Template DNA	-	5 uL

Table 7 Cycling conditions for NGS library amplification

Step	Temperature	Time	Number of Cycles
Initial denaturation	94C	3 min	1
Denaturation	94C	30 sec	35 cycles
Annealing	60C	40 sec	
Extension	72C	90 sec	
Final extension	72C	10 min	1
Cooling	10C	HOLD	1

The resulting amplicon was cleaned and pooled in 96 wells using SequalPrep normalization plate kits (Invitrogen) and AMPure XP beads (Beckman Coulter) both as per manufacturer's protocol. Ratios of beads used varied from 0.6 to 1.6:1, and both plate and columnar magnets were used, both prior to and after SequalPrep normalization (See section 10.2 and 14.3 for details). The pooled 96 well plates were diluted to a 2nM final library after quantification using a Qubit 2.0 (Life technologies), a KAPA library quantification kit (KAPA Biosystems), and an Agilent high sensitivity

DNA kit (hsDNA) using a Bioanalyser or an Agilent 2200 TapeStation Instrument. The library was loaded onto a MiSeq (Illumina) as per manufacturer's protocol for 500 cycle V2 kits with the addition of custom sequencing primers using a 10-50% PhiX spike.

Table 8 Custom Sequencing Primers

Read 1 sequencing primer			
Padding	785F primer		
TACCGGGACTTA	GGATTAGATACCCBRGTAGTC		R1Seq
Read 2 sequencing primer			
Padding	1175R primer		
AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC		R2Seq
Index 2 sequencing primer			
Reverse complement of 1175R primer	Reverse complement of Read 2 padding		
GAGGAAGGHGGGGAYGACGT	TAAAACGTGTT		In2Seq

The resulting FASTA files were analysed using QIIME[198] (Section 19.3, Appendix B). Paired-end sequenced reads from each MiSeq run were merged using FLASH[217] and demultiplexed, pooled and assigned OTUs using QIIME v1.8.0[1]. OTUs were clustered at 97% similarity and assigned taxonomy against the Greengenes database (v.13.5)[2]. OTUs occurring at a count fraction of <0.001 were filtered out (minimum count fraction or MCF), and a single random rarefaction carried out to produce an even sampling depth across all samples (1000 sequences per sample). Further analyses on both rarefied and unrarefied results were carried out in R Studio (v.0.99.896) using Phyloseq (v.1.10.0). In a secondary analysis OTUs seen in experimental negative controls were removed from both rarefied and unrarefied results in Phyloseq.

9.8 Assessment of Gastrointestinal Tract Injury and Anthropometry

All samples were assayed on first thaw using the commercial ELISA kit (Hycult) with plasma diluted 1 in 10 as per protocol. The measurable concentration range of the assay was 20-5000 pg/mL. 100µL of standards and diluted samples were incubated for an hour in supplied microwell plates at room temperature. They were washed 4 times using provided wash buffer and 100µL of diluted tracer added to each well, incubated for an hour and the wash repeated. 100µL of diluted streptavidin peroxidase was then added followed by another hour incubation and a further wash. 100µL tetramethylbenzidine substrate was then added, the plate covered with tin foil and incubated for half an hour before 100µL stop solution was added. The plate was read using plate reader (Infinite F2000, Tecan) at 450nm within 30 minutes. Standards were run in duplicate and samples in singlicate owing to limited volume of sample available. The mean absorbance for each set of standards and samples was calculated, and samples giving absorbance >15% of the mean for that set were considered suspect and retested, as were samples which were undetectable. If the mean absorbance of the blank/zero standard was >0.4, the run was repeated. Absorbance of the blank/zero standard was subtracted from the absorbance of the samples prior to standard curve fitting. A standard curve was fitted using Microsoft Excel® for Mac 2011 (version 14.6.6), and R^2 >0.99 were considered acceptable.

Anthropometric measurements were compared with molecular test results over time as a proxy markers of intestinal dysfunction[3,4].

9.9 Comparison of Clinical Events to Molecular Test Results

Clinical events were collated from a) adverse event reports b) reports of admission completed by trial nurses if a child needed hospital admission c) microbiological investigation performed by trial team. A comparison was made with children who had a high result from a molecular test (qPCR/16S broad range PCR result >1000 CFU Equivalents/PCR reaction, or successful sequencing with NGS) within four weeks of the event, those who did not have a high test result or successfully sequenced with NGS and had a study sample available within four weeks of the event, and those who had a clinical event but had no study sample available within four weeks of the event. The comparison investigated the frequency and types of event recorded in each

group. The molecular test results at the nearest time point for children who had either a confirmed urine or blood culture positive were also documented, even if these fell below the threshold of a PCR result >1000 CFU Equivalents/PCR reaction.

9.10 Statistical Analysis

Statistical analysis was carried out in R Studio (v.0.99.896). Descriptive analysis included simple tabulations and percentages alongside chi-squared and Fisher's exact test for binary data (depending on sample numbers available), and Wilcoxon paired tests and Mann-Whitney-U tests for numeric data. Correlation calculations were carried out to calculate Spearman's rho. Comparisons were made between ART groups (experienced, naïve, naïve-matched controls, experienced-matched controls, Gulu (rural group, ART-naïve) and experimental controls where appropriate. Comparisons were also made between the results from different assays (e.g. between NGS and qPCR results). For analyzing the NGS results, principal coordinates analysis was carried out in R Studio (v.0.99.896) using Phyloseq (v.1.10.0) and the Unifrac method.

10 Chapter 10 Assay Development Results

10.1 QPCR Assay Development

10.1.1 Identifying Target Organisms

The literature review prioritised studies that documented the prevalence of genera affected by HIV infection when studying the gut microbiome[5-7], or when investigating microbial DNA in the bloodstream of those infected with HIV[8,9]. These studies identified *Enterobacteriales*, *Lactobacillales*, *Pseudomonandales*, *Burkholderiales* and *Bacillales* as key genera. In particular, *Bifidobacteria* and *Lactobacillales* numbers appeared to be adversely affected while the proportion of *Pseudomonandales* increased in adults with HIV infection versus controls[10].

Few studies were identified where detected bacterial DNA in the blood stream of people with HIV was sequenced, so the choice of targets was further informed by available literature documenting sequenced bacterial DNA identified from stool samples. Ellis *et al.* investigated differences in faecal microbiota in HIV, comparing order abundances with patient clinical status and evidence of immune activation in the gut itself[5]. Microbial DNA was extracted from faeces and subjected to 16S rDNA qPCR in order to quantify pan-bacterial loads and relative abundances of *Enterobacteriales*, *Bacteroidales*, and *Clostridiales*. Increased *Enterobacteriales* detected in stool was positively associated with duodenal CD4 T cell depletion[5].

Considering evidence from humans uninfected with HIV, but affected by malnutrition, Smith *et al.* investigated differences in gut microbiota from Malawian twin pairs, some healthy and some discordant for kwashiorkor. DNA was obtained from fecal samples and subjected to multiplex shotgun pyrosequencing.[11] Similarly, they found predominance of *Bacteroides*, *Bifidobacterium*, *Clostridiales*, *Escherichia coli*, *Ruminococcus*, *Lactobacillus*, and *Enterococcus*.[11] Significant changes were seen in the microbiota of gnotobiotic mice with associated weight loss after receiving fecal

microbiotal transplant from twins discordant for kwashiorkor, and on exposure to Malawian diet versus ready to use therapeutic food[11].

Data from healthy humans included sequencing of 16S rDNA clones obtained from 10 cycle PCR from stool samples from a healthy 40 year old man. This demonstrated 95% OTUs obtained were from 3 groups: the *Bacteroides* group, the *Clostridium coccoides* group, and the *Clostridium leptum* subgroup.[12] Changes in gut microbiota occur with age: *Bifidobacterium*, *Streptococcus, Lactococcus,* and *Lactobacillus* dominate the infant gut microbiota, with lower numbers of *Bacteroidetes, Firmicutes*, and *Archaea*[13]. A further study investigated a more broad population from differing geographic locations. Similar results were obtained, although *Enterococcaceae* were overrepresented in Malawian versus Amerindian infant microbiota[13].

Different bacterial loads and profiles within different areas of the gut were taken into account. In a study taking mucosal biopsies at different locations in the gut and comparing bacterial diversity of PCR-amplified 16S rDNA clone libraries, the jejunal samples were dominated by species related to *Streptococcus* (67%) while more distally, the sequences from the distal ileum, colon and rectum were dominated by *Bacteroidetes* (27-49%) and *Clostridium* clusters XIVa (20-34%) and IV (7-13%).[14]

To summarise, the following bacteria were identified as important potential targets qPCR targets: *Lactobacillus* (genus), *Bifidobacterium* (genus), *Clostridiales* (order), Enterobacteriaceae (family), *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Streptococcus pyogenes*, and *Fusobacterium*. Assays for the following were already in use prior to the study: *Enterobacteriaceae* family (target gene *DnaK*)[15], *Staphylococcus aureus* (target gene *CoA*)[16], *Streptococcus pyogenes* (Target gene *csrR*)[15], *Fusobacterium* (16S rDNA)[17], *Staphylococcus* species (target gene *tuf*)[15]. Assays for *Lactobacillus* (genus), *Bifidobacterium* (genus), *Clostridiales* (order) were lacking.

10.1.2 Identification of Target Genes for Amplification

Bifidobacterium

For *Bifidobacterium*, targets considered included the unique fructose-6-phosphate phosphoketolase pathway used to ferment carbohydrates, genes involved in the stress response to bile, temperature and acid[18]; colonisation genes such as the type IVb tight adherence (Tad) pilus-encoding gene cluster *tad2003*[19]; and immunomodulation genes that encode serpin-like protease inhibitors that interact with human neutrophils and pancreatic elastases[20]. After checking ubiquity among Bifidobacterial species and sequence homology, the most promising target was the *xfp* gene which encodes the D-xylulose 5-phosphate / D-fructose 6-phosphate (X5P / F6P) phosphoketolase (Xfp), the key enzyme of the F6P-phosphoketolase pathway in bifidobacteria[21,22]. Phosphoketolases (EC 4.1.2.9, EC 4.1.2.22) are thiamine diphosphate (ThDP)-dependent key enzymes of the pentose phosphate pathway lactic acid bacteria and of the d-fructose 6-phosphate (F6P) shunt of Bifidobacteria[23].

Lactobacillus

Acid, osmotic and bile salt stress resistance genes were screened along with cell surface adhesion genes (which were too variable) [24,25]. Metabolic genes such as those involved in amino acid synthesis were also screened[26], along with quorum sensing genes, but given that these had the potential for horizontal transfer they were judged unlikely to be sufficiently specific[27]. The most promising gene was a two component regulatory system LBA1524HPK-LBA1525RR 2CRS associated with resistance to acid stress[26]. The HPK gene had 36% identity with HPK in *lis*RK in *Listeria monocytogenes*.[28] This gene was subjected to the screening processes described below with design of a primer-probe set. However, on testing the set with isolates of *Lactobacillus reuteri, Lactobacillus rhamnose, Lactobacillus casei, Lactobacillus plantarum* and *Lactobacillus gasseri* the reaction failed and this primer set was abandoned. After further literature review and discussion with experts in the field, it was decided to target a ubiquitous house keeping gene *groEL* (coding heat shock protein 60: *hsp*60)[29](personal communication Dr Makarova, 4.10.13). Sequences were collated

from key branches of *Lactobacillus* families as well as possible overlaps (with *Streptococcus pneumoniae, S. agalacticae, S. pyogenes, S. thermophiles* and *Lactococcus* spp) to ensure specificity[29]. Owing to sequence divergence between families, the primer set were designed with one forward primer, 2 reverse and 2 probes to cover the families including *Lactobacillus casei* and *Lactobacillus acidophilus* most effectively.

Clostridiales

Potential conserved genes screened included spore forming genes[30-32] and butyrate producing genes as butyrate appears to have a role in moderating the inflammatory process of gut disease (for example Crohn's disease and Ulcerative Colitis) and as an energy source for the colonic epithelium[33-35] both of which were found to be insufficiently conserved. Primers have previously been designed for the butyryl-CoA CoA transferase but when compared in a multiple sequence alignment using MegAlign (Lasergene10, DNAStar) they were very degenerate and also had a relatively low sensitivity when tested using isolates- which is a much more challenging for samples likely to have a low bacterial load (such as plasma) than for assessing faecal samples[36]. Primers targeting the 16S rDNA gene were assessed and trialed in the absence of another well-conserved gene. This is unsurprising given the size and diversity of the *Clostridiales* family as well as frequent taxonomic reassignments.

10.1.3 Designing Primer and Probe Sets

Primer probe sets were designed for *Bifidobacterium*, *Lactobacillus* and *Clostridiales* as per section 9.4.3.

Clostridiales

Widely used primer sets for *Clostridiales* have been targeted to a conserved area of the 16S rDNA gene, with two sets developed to target two broad groups of Clostridia

using a SYBRgreen assay. These were the *Clostridium leptum* group (including *Eubacterium siraeum, Clostridium cellulosi, Clostridium orbiscindens, Clostridium sporosphaeroides* and *Eubacterium desmolans*) and the Clostridium coccoides group (including *Ruminococcus lactaris, Butyrivibrio fibrisolvens, Clostridium clostridioforme,* and *Clostridium nexile*)[37,38 39]*.* When comparing the primer sets from Matsuki *et al.* and Rintilla *et al.* using MegAlign they were found to target overlapping regions so only the Matsuki *et al.* sets were trialed as they appeared to target more conserved regions.

10.1.4 Testing Primers against Target Strains

Bifidobacterium

These primers were tested using the reaction conditions described in Section 9.4.4 (reaction 1), using 1.0 µL of forward and reverse primers. The limit of detection for *Bifidobacterium longum, Bifidobacterium adolescentis* and *Bifidobacterium infantis* was 0.5 CFU/µL. No *Lactobacillus, Clostridiales* or *Streptococcal* species were detected, nor *Escherichia coli* or *Staphylococcus aureus*.

Lactobacillus

These primers were tested using the reaction conditions described in Section 9.4.4 (reaction 2), using 1.0 µL of forward and 0.5µL of reverse primers. Cycling conditions as per section 9.4.4. Owing to reduced sensitivity for *L. reuteri* with this mix, the assay was re-run using just the acidophilus probe and reverse primer with resultant improved sensitivity.

Table 9 Sensitivity of assay for detecting different *Lactobacilli*

Species	Limit of detection (cfu/PCR reaction)
L Casei	1.3-13
L Rhamnose GG	1.4-14
L Reuteri (using R and P acidophilus)	5.1-51
L plantarum	Undetected

Bifidobacterium, Clostridiales or *Streptococcal* species were not detected, nor *Escherichia coli* or *Staphylococcus aureus*.

Clostridiales

Reaction conditions as described in Section 9.4.3 (reaction 3). Isolates included were: *Clostridium butyricum, Clostridium sordelli, Clostridiaceae* bacterium, *Clostridium orbiscindens, Lachnospiraceae, Clostridium clostridioforme, Clostridium aldenense, Clostridium hathewayi, Clostridium citroniae* and *Streptococcus mitis.* The *C. coccoides* primer set detected *C. butyricum* at 10-100 CFU/PCR reaction but none of the other strains. The *C. leptum* set detected *Streptococcus mitis* at the same frequency as any of the Clostridial strains. At this stage, given the lack of sensitivity and specificity of this primer probe mix and the high chance of futility given the diversity of the *Clostridiales* genus, it was decided to focus on broad range techniques for detection rather than specific.

10.1.5 Testing Experimental Samples

Two primer sets targeted entire bacterial families (*Enterobacteriaceae* and *Staphylococcae*) and so were more vulnerable to contamination than those which were species specific, as demonstrated both by Tann *et al*. and in this study (negative controls and negative extract samples crossing the cycle threshold between 35-40 cycles)[15]. Therefore a lower CT cut off was set for these assays. In order to draw

comparisons between assays, each of which is subject to some degree of variability as well as distinct sensitivities and specificities, a schema was drawn up to enable comparison between these heterogenous assays. Where samples were positive for more than one sample, the codes were summed to enable crude but standardised comparison between results. See Section 11 for full results.

Table 10 Scheme for interpretation and comparison of PCR results

Normal Cut off PCRs	Low Cut off PCRs (Ent/TUF)	CFU equivalent	Ranking	Coded
>40.1	>35.1	0	Negative	0
37.0-40	32.0-35	1-10	Very low	1
34.0-36.9	29-31.9	10.1-100	low	2
31.0-33.9	26-28.9	100.1-1000	moderate	3
28.0-30.9	22.0-25.9	1000.1-10000	high	4
<28	<22	>10000.1	Very high	5

10.2 Next Generation Sequencing Assay Development

Initial assay development involved using blood samples from 5 healthy volunteers. These were used in triplicate- once as whole EDTA blood, once as EDTA plasma and once as a cell pellet in order to compare results depending on sample constituents, alongside controls as per section 9.7. Samples were extracted as per section 9.3. Library preparation, normalisation, sequencing and analysis were carried out as per section 9.7.

10.2.1 Contamination of Barcoded Primers

On analyzing the results of the initial assay development run, it was noted that the most prevalent organism was *Sphingomonas,* with nearly 2 million reads. It appeared in every sample and was the most abundant sequence found in the greater part. The next most prevalent was *Porphyrymonas gingivalis* with a read count of 400000, limited to only two samples previously demonstrated to have contained this organism. Having ruled out all other reagents as a source of contamination, the

barcoded primer sets were tested by using them as template in a standard 16S rDNA PCR and sequencing any resulting amplicons as per sections 9.5 and 9.6. The resulting sequences were found to be >98% homologous with *Sphingomonas spp.* The sequence obtained was then aligned using MegAlign with sequences from the NGS run and found to be identical.

10.2.2 Primer Dimer

After sourcing new primers, sample preparation of study samples was carried out as per Sections 9.3 and 9.7 the same method was used with samples from this study. However there were problems with discrepant quantification results post-pooling and cleaning using the Qubit such that two runs failed due to over- and under-loading of the MiSeq (the library although appearing to be 10pM was either too concentrated or too dilute to successfully sequence). To overcome this quantification with a KAPA library quantification kit (KAPA Biosystems) was used at several stages during the normalizing and pooling, and directly before loading the library. Despite using the KAPA quantification kit, difficulties were still encountered in determining the correct loading concentration. The use of the Agilent high sensitivity DNA kit (hsDNA) on the Bioanalyser revealed the presence of a high concentration of primer dimer (Figure 16) which explained the previous inaccuracy in library quantification. Comparisons of the relative results using the KAPA quantification kit, the Qubit and the Agilent hsDNA kit on four 96 well plates post normalization and cleaning using AMPure beads are in Table 11.

Table 11 Comparison of Qubit and KAPA results for library quantification

	Qubit	KAPA Quantification kit
	Conc (nM)	Conc (nM)
Plate 1	155741	28633
Plate 2	196185	29893
Plate 3	144875	27480
Plate 4	130388	

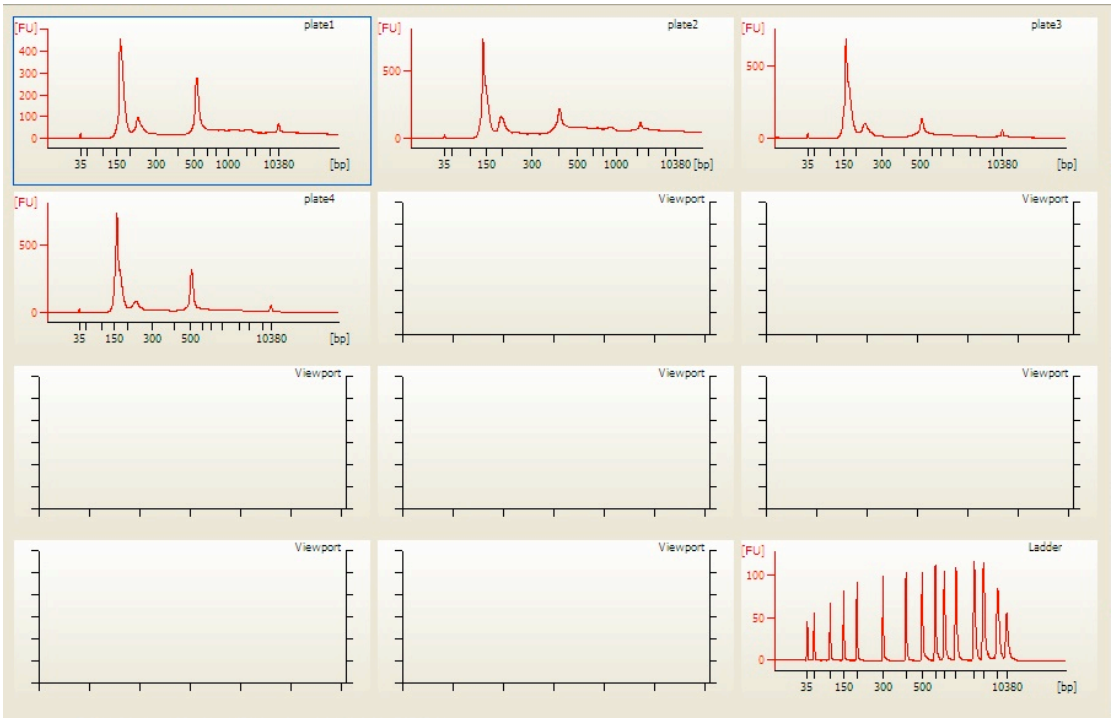


Figure 16 Agilent hsDNA kit results demonstrating primer dimer in four plates of library preparation

A large primer dimer peak can be seen at 150bp in all 4 plates at a much higher concentration than the amplicon peak at 502bp

The plates in Figure 16 were subjected to 3 repeat size selections using AMPure XP beads at a ratio of 0.8:1 beads to sample (Beckman Coulter) to reduce the relative concentration of primer dimer to amplicon. A ratio of 0.6:1 beads to sample was trialled to see if further reduction of the primer dimer could be achieved but this was found to give similar results with a more pronounced loss of amplicon (Figure 17).

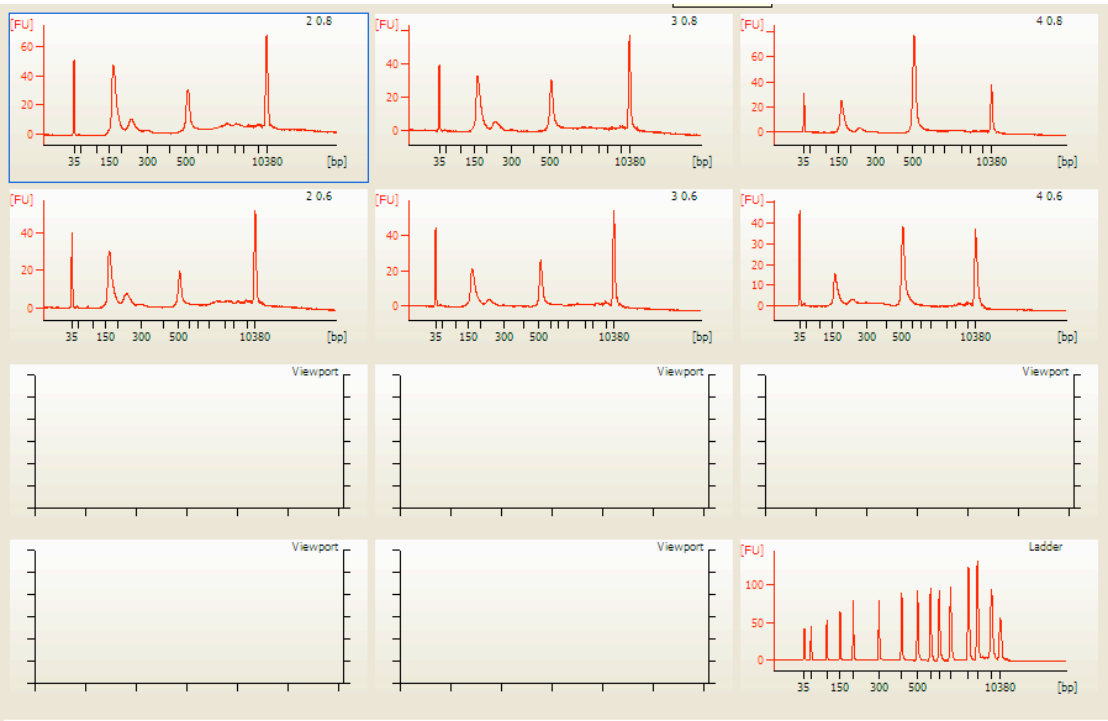


Figure 17 Comparison of 0.8:1 (top panel) and 0.6:1 (second panel) AMPure beads to sample ratio in reducing primer dimers in plate numbers 2, 3 and 4

After pooling, attempts to run this library both with a 10% and a 50% PhiX spike failed. Repeating the pooling of the four plates and analysing on an Agilent 2200 TapeStation Instrument gave a peak molarity of 385 pmol/L for the primer dimer, and 108 pmol/L for the amplicon. At this stage, this library was abandoned as 80% of the potential sequencing capacity would be lost even if the sequencing was successful.

The library amplification was repeated in an attempt to reduce the primer dimer generation and remove it more effectively. Firstly, only samples with CFUs/PCR reaction>100 using the qPCR were selected to maximize the potential for successful amplification. Halving the concentration of primer at the library amplification stage using controls was found to give satisfactory amplification. Therefore the amplification was performed using the primers at a concentration of 0.2pmol/µL. The post amplification normalization was altered so size selection using AMPure beads (ratio 0.8:1 beads to sample) and a plate magnet so each sample was subjected to size selection prior to pooling. This was followed by using the SeqPrep normalization kit, and several further size selections using AMPure beads, initially at a ratio of 1.8:1

and then 0.8:1 beads to sample. The resultant amplicon to primer dimer peaks are shown in Figure 18.

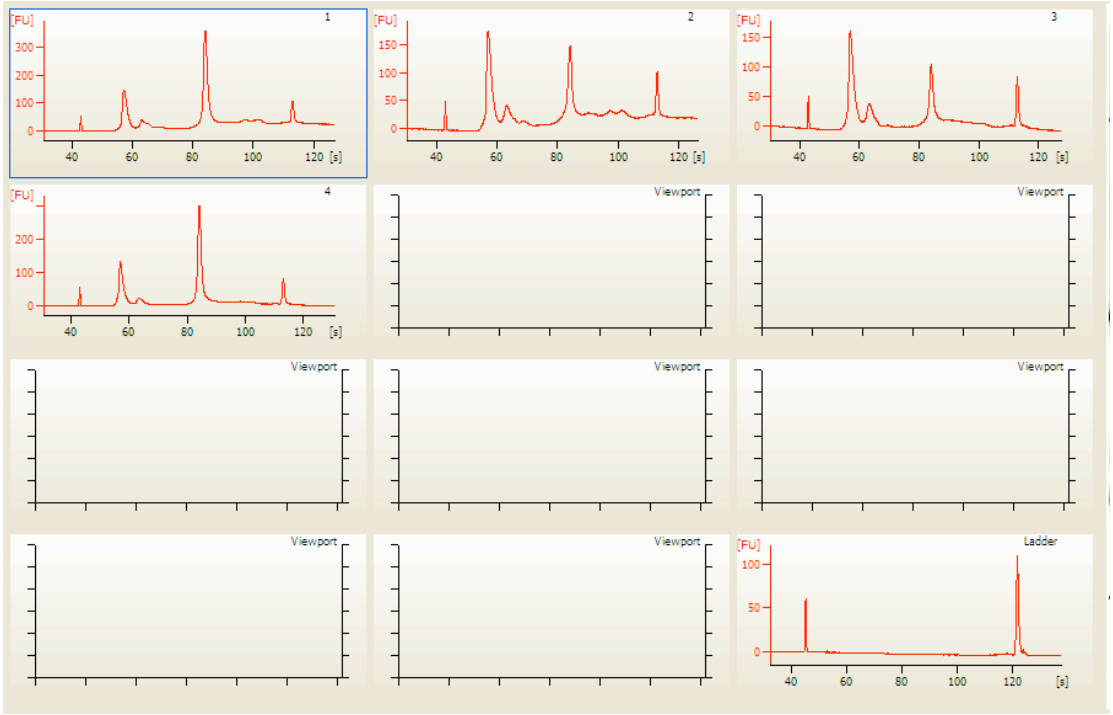


Figure 18 Amplicon versus primer dimer peaks using Agilent Bioanalyser in 4 plates of pooled library post-size selection and altering method of normalisation

After pooling of all 4 plates, the concentrations of amplicon and primer dimer were 740 pmol/L and 949 pmol/L giving a ratio of amplicon to primer dimer of 0.78. This library failed to sequence. Therefore, plate 1 (top left in Figure 18) which had the highest concentration of amplicon (1323 pmol/L, primer dimer concentration 408 pmol/L, ratio 3.24) was chosen to sequence in isolation, abandoning the other 3 plates which contained a much lower proportion of amplicon. This plate (which mostly contained cell pellet samples) sequenced successfully generating 8.2M reads prior to analysis.

Four further 96 well plates of samples were amplified using the same protocol and a similar pattern emerged (Figure 19)

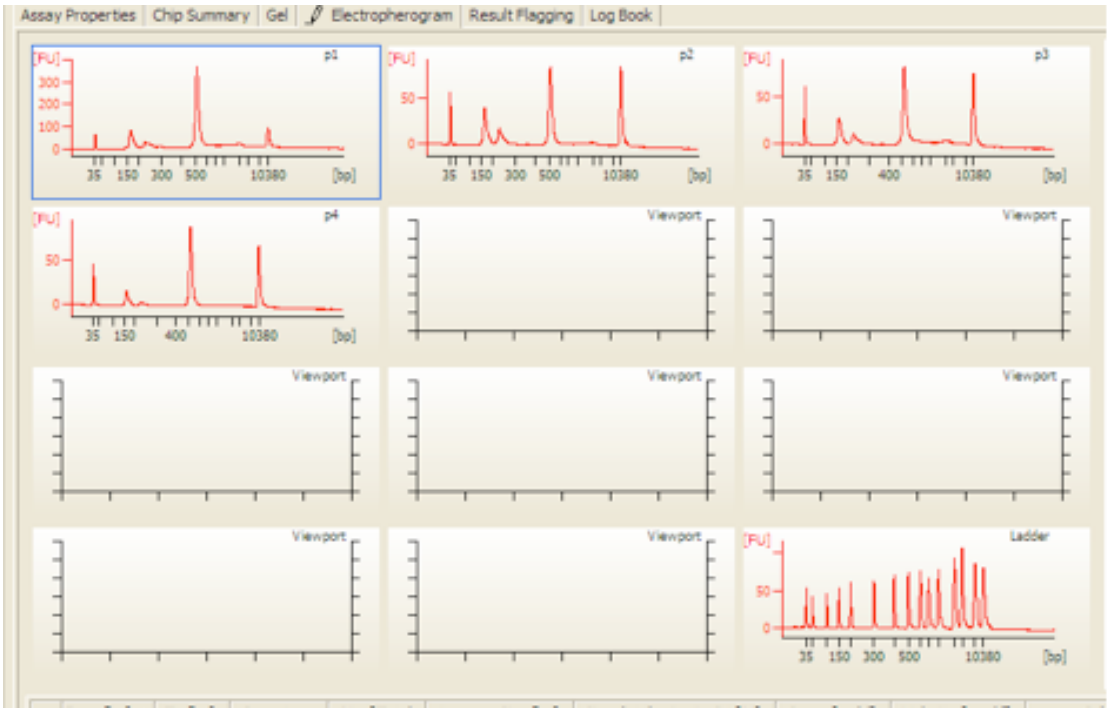


Figure 19 Results of amplification of four further 96 well plates labelled p1-p4 using Agilent Bioanalyser demonstrating primer dimer and amplicon peaks

Table 12 Comparison of Primer dimer to amplicon concentrations and Qubit results for second library amplification

Name	Primer dimer (pmol/L)	Amplicon (pmol/L)	Qubit nM	Ratio Amplicon: Primer dimer
Plate 1	4871	2721	7.24	0.56
Plate 2	2466	669	2.44	0.27
Plate 3	2072	780	2.92	0.38
Plate 4	965	734	1.65	0.76

Repeated size selections were performed on Plate 1 (Table 13) to attempt to increase the ratio further.

Table 13 Table demonstrating relative reduction in both primer dimer and amplicon concentration using repeat size selection with AMPure beads

Number of 0.8 size selections	Primer Dimer pmol/L	Ratio reduction by size selection	Amplicon pmol/L	Ratio reduction by size selection	Ratio Amplicon: primer dimer
2	9318		3283		0.35
3	4871	0.52	2721	0.83	0.56
4	2220	0.46	1830	0.67	0.82
5	1420	0.64	1010	0.55	0.71
6	743	0.52	719	0.71	0.97
7	67	0.09	390	0.54	5.82

This plate was sequenced successfully generating 4.8M reads prior to analysis.

One further attempt was made to improve the output from NGS. The barcoded primers were redesigned using IDT's oligoanalyser programme (Integrated DNA Technologies)[40] to have a higher melting temperature (Tm). The padding sequence length was increased and the barcodes were exchanged for Nextera™ (Illumina) barcodes which are 8 bp as opposed to 12 bp[41], which increased the Tm from ~62°C to ~68°C (Appendix B Tables of barcoded primer sequences for NGS, Table 40).

The 94 samples with the highest CFUs/PCR reaction (as per 16S rDNA PCR using the SYBRgreen™ assay) were amplified with the redesigned primers as per conditions above, and immediately post library amplification were quantified using the Agilent 2200 TapeStation™. Only 6 samples had an amplicon peak. The relative size of the amplicon to the primer dimer peak was quantified as per Table 14.

Table 14 Comparison of Relative Primer Dimer to Amplicon Concentrations using redesigned primers as assessed by Agilent 2200 TapeStation

Sample Identifier	Molarity (pmol/L)		% total	
	primer dimer	amplicon	primer dimer	amplicon
C1	14000	12300	17	54
D2	38200	2360	56	12
C3	37500	3010	54	15
C9	74600	1710	65	5
C10	38000	2960	35	10
B11	17000	14700	18	56

Only 2 samples contained amplicon in sufficient concentrations that post normalization and clean up they were likely to sequence successfully. These two samples also generated high numbers of OTUs using the previous primer set. Therefore minimal increase in data is likely to be generated by proceeding with further library amplification with the redesigned primers.

10.3 Assay Development Discussion

10.3.1 Quantitative PCR Assay Development

The key assay development issues with regards to qPCR primer design were the challenges of designing primers for a family such as the *Clostridiales* which have a high degree of diversity between species. Even using previously published primers, a high degree of overlap was found between *Streptococcus mitis* and Clostridial species, meaning that primers targeting this family were not sufficiently specific to amplify only Clostridial species. As the main advantage of qPCR assays is their specificity, we took the decision to focus on broad range assays, such as broad range 16S rDNA PCR with NGS to detect these diverse families of bacteria. Two other primer

sets were used to detect whole families (*Enterobacteriaceae* and *Staphylococcae*), and these assays are therefore more vulnerable to contamination as demonstrated here and in previous studies[15]. A lower CT value was used to determine positive samples with these assays than in the other species specific assays, and a scheme was drawn up to enable comparison between these assays. As the assays are heterogenous and will have some degree of variability, it is important that results such as precise CFUs are not over-interpreted. The developed schema (Table 10) aims to enable broad comparison between assays while accounting for that variation. When samples had positive results for more than one assay, the code derived was summed to enable a crude comparison between assays. This was for hypothesis generation e.g. investigating whether samples positive for several different species also had a higher result using a broad range 16S rDNA PCR. This approach is somewhat crude and has its limitations, but does allow a basic standardised comparison between assays.

10.3.2 NGS Assay Development

A number of issues were encountered in the development of the NGS assay. First was the discovery that the barcoded primers, obtained from a commercial source, were contaminated with *Sphingomonas sp.* This waterborne bacterial species is only rarely found in clinical samples, usually in immunocompromised individuals. The *Sphingomonas sp.* detected during the first NGS run was identical to that identified by Sanger sequencing an amplicon derived from using the primers themselves as a template, which appears to demonstrate that the primers themselves were the source of the contamination. On reviewing previous runs performed in the same laboratory using this primer set, *Sphingomonas* sp. was also noted, albeit proportionally lower as previously tested samples had a higher *a priori* bacterial load. The risk of reagent contamination using NGS, particularly from low biomass samples, has been noted by others, with various techniques suggested to remove contamination[42,43]. As yet there is no clear consensus about how best to avoid contamination, although UV irradiation of reagents such as water should be used whenever possible[43].

The second major issue was of the difficulties encountered when attempting to accurately quantify the pooled library prior to MiSeq loading. Initially, a KAPA quantification kit was used. The advantage of the KAPA method is that it aims to calculate the concentration of the amplified library only rather than all of the DNA present in the sample, allowing increased accuracy. However, the MiSeq loading concentration is calibrated using the Qubit to determine total DNA content of the sample, rather than just the amplified library, as the assumption is that that the amount of other DNA present will be minimal. In this case, as the KAPA method targets the Illumina® adaptor sequences, it may be that primer dimer was also amplified, which would interfere with the final calculation of DNA concentration.

It was necessary to use the Agilent hsDNA kit to identify the primer dimers which contributed to the overall DNA concentration and rendered quantification by both Qubit™ and KAPA™ methods inaccurate. The primer dimers possibly interfered with binding sites on the flowcell, in particular as the primer dimer concentration was so much higher than that of the amplicon. This was seen even when the spike of PhiX was increased to 50%: cluster generation on the flow cell was insufficient likely due to primer dimer interference.

To overcome this, first, I prioritised using samples which contained higher bacterial loads on screening with a broad range 16S rDNA PCR. Secondly, I used half the volume of primers in the initial library preparation. Thirdly, I introduced a number of steps to remove the primer dimer during the normalisation process. These included using a plate magnet with AMPure XP beads prior to sample pooling with the SequalPrep kit, and numerous further clean ups of pooled samples using the AMPure XP beads post pooling alongside quantification with the Agilent hsDNA assay on a Bioanalyser and Tapestation™ instrument. As the amplicon was already at a low concentration, a threshold was reached whereby further clean-ups would proportionally reduce the amplicon more than the primer dimer and sequencing would be likely to fail (Table 12). This threshold differed between samples and pooled plates dependent on the success of the initial library amplification so in the future,

performing individual hsDNA assays using the Tapestation™ instrument prior to library clean up and pooling is likely to be the most productive way of identifying samples that are likely to be sequenced successfully. The final attempt to improve output was to amend the primers and increase their melting temperatures (Tm). A higher Tm will be better suited to the high temperatures of sequencing and should improve the performance of the sequencing primers[41,44]. The final methods used for broad range 16S rDNA PCR with NGS are described in the following chapters.

11 Chapter 11 Description of Study Cohort

11.1 Background

This study was carried out using samples from the CHAPAS-3 trial (ISRCTN69078957) as described in section 9.1. In order to fully interpret the biological impact of the results described in subsequent chapters, it is necessary to first describe the children enrolled in more detail.

11.2 Study Population

See section 9.1

11.3 Methods

Statistical analysis was carried out as per section 9.9

11.4 Results

CHAPAS-3 enrolled 480 children from Uganda and Zambia, randomising 156 to d4T, 158 to AZT and 164 to ABC (2 excluded due to randomisation error). The overall results showed good clinical, immunological and virological results in all three groups with low and comparable toxicity rates in each arm[45]. There were 19 deaths, all in ART-naïve children, 9 of which occurred in the first 12 weeks of ART.

The cohort for this study included 305 children altogether: 119 ART naïve, 22 experienced, 89 controls age matched to ART-naïve children, and 20 controls age matched to ART-experienced children all from the urban (Joint Clinical Research Centre (JCRC) Kampala site, and 55 ART-naïve children from the rural Gulu site. At enrolment, the ART-experienced group had spent a median duration of 4 years on ART (IQR 2.6-4.3). Demographics, baseline CD4 T cell counts and percentages, viral loads, percentage of HIV-infected children in each group with a fully suppressed viral load at week 96 of follow up and anthropometric data are presented in

Table 15. ART naïve children tended to be younger, have lower CD4 T cell counts and percentages at baseline, and lower anthropometric markers at baseline than ART-experienced children (Figure 20, Figure 22). They also experienced greater improvement over time in terms of CD4 T cell count/percentage and anthropometric markers over the course of the study than the ART-experienced group (Figure 20, Figure 23). In comparison with controls, the ART naïve group were younger than the age-matched controls (median age 2.8 years, 95% CI 2.6-3 versus median age 3.3, 95% CI 3-3.6, p<0.001), but the ART-experienced group were similar in age (median 6.5 years, 95% CI 5.6-7.3 versus 6.3, 95% CI 5.3-7.3, p=0.12). Both HIV-infected groups had lower CD4 T cell percentages than controls, though the ART-naïve group more markedly so (Figure 20). The ART naïve group had lower anthropometric markers than controls, but the ART-experienced group were more comparable with their age-matched peers (Figure 22).

The Gulu group were slightly, though not significantly, younger than the ART naïve children from JCRC, and had higher baseline CD4 T cell counts with borderline lower viral loads at baseline (Figure 20). They had the lowest rate of viral suppression at 96 weeks (67% below 100 copies/ml) although this did not reach statistical significance in comparison with the other two groups (Figure 21). In terms of anthropometry, they were broadly similar in baseline characteristics and improvement over time to the JCRC ART-naïve group (Figure 22, Figure 23).

There were 6 deaths in the cohort, all in ART naïve children (3 from JCRC, 3 from Gulu). Age, cause of death and duration of follow up at date of death are in Table 17. Two died early within 6 weeks of enrolment, and the remaining four had at least 24 weeks follow up. Four patients (3%) were lost to follow up by week 72 from the ART-naïve group. Of these four, one only had baseline samples, while the other three had both baseline and week 12 samples. Three were lost to follow up (5%) from the Gulu (ART-naïve) group, one within 6 weeks of randomisation and the other two between week 12 and week 72. All children from the ART-experienced group had samples available at week 72.

Table 15 Demographics, CD4 T cell counts, viral loads and anthropometric data compared by ART group

ART group	Naïve	Experienced	P value naïve versus experienced	Naïve Controls	Experienced Controls	P value cases versus controls	Gulu	P value (Gulu vs naïve cases)
Number	119	22		89	20		55	
Male Sex (%)	58 (49)	10 (45)	0.82	40 (44)	12 (60)	0.65 (naïve) 0.56 (exp)	27(49)	0.99
Age (years)(median, IQR)	2.8 1.7-4	6.45 5.9-9.2	**<0.001**	3.3 2.4-4.4	6.3 5.7-8.9	**0.006** (naïve) 0.89 (exp)	2 1.7-3.5	0.12
Age range	0.3-5	5.1-11.2		0.6-7.4	5.1-11.8		0.6-4.8	
Baseline CD4 T cell count (mm^3) (median, IQR)	925 637-1451	1188 927-1813	**0.02**	1333 1036-1707	1010 856-1343	**<0.001** (naïve) 0.11 (exp)	1031 674-1602	0.51
Baseline CD4 T cell percentage (median, IQR)	20 14-24	35 31-39	**<0.001**	38 34-43	40 35-45	**<0.001** (naïve) 0.06 (exp)	30 22-41	**<0.001**
Baseline Viral Load (copies/ml) (median, IQR)	401200 161300-1021000	<100	NA	NA	NA	NA	267400 130100-746300	0.047
CD4 T cell count change baseline-Week 96 (mm^3) (median, IQR)	314 -41-657	-157 -438-59	**<0.0001**	NA	NA	NA	NA	NA
CD4 T cell % change baseline-Week 96 (mm^3) (median, IQR)	17 12-22	3 0-6	**<0.0001**	NA	NA	NA	NA	NA
Suppressed VL at Week 96 (%) 95%CI	84/110 (76%) 67-83%	19/21 (91%) 70-99%	0.243	NA	NA	NA	30/45 (67%) 51-80%	0.23
Weight for age z score Baseline (median, IQR)	-2.0 -3.2- -0.9	-1.5 -2.1- -0.4	0.06	-0.6 -1.2- 0.1	-1 -2.1- -0.1	**<0.0001** (naïve) 0.54 (exp)	-1.8 -3.1- -1.0	0.85
Height for age z score Baseline (median, IQR)	-2.5 -3.5- -1.3	-1.5 -2.1- -1.1	0**.02**	-0.8 -2.1- 0	-1.3 -2.5- -0.3	**<0.0001** (naïve) 0.52 (exp)	-2.4 -4.6- -1.6	0.24

BMI for age z score Baseline (median, IQR)	-0.5 -1.2-0.5	-0.3 -1- -0.5	0.73	-0.1 -0.7- 0.8	0 -0.9- -0.4	0.01 (naïve) 0.48 (exp)	0 -0.9- 1	0.07
Weight for age z score change Baseline- W96 (median, IQR)	1 0.2-1.9	-0.1 -0.2- 0.0	**<0.0001**	NA	NA	NA	0.6 0.18-1.3	0.23
Height for age z score change Baseline W96 (median, IQR)	0.8 0.2-1.3	-0.1 -0.3- -1.4	**<0.001**	NA	NA	NA	0.8 -0.2- 1.8	0.97
BMI for age z score change Baseline W96 (median, IQR)	0.6 -0.1- 1.6	-0.1 -0.4-0.1	**0.001**	NA	NA	NA	0.2 -1.1- 1.1	0.08

The ART randomisations for the HIV infected groups are shown in Table 16.

Table 16 ART randomisation for HIV infected children

Arm	Received	Total	JCRC (urban) ART-Naïve	JCRC (urban) ART-Experienced	Gulu (rural), ART naive
1	D4T	66	42	7	17
2	AZT	57	35	7	15
3	ABC	73	42	8	23
4	Controls	109	0	0	0
NNRTI 1	EFV	48	46	2	
NNRTI 2	NVP	93	73	20	

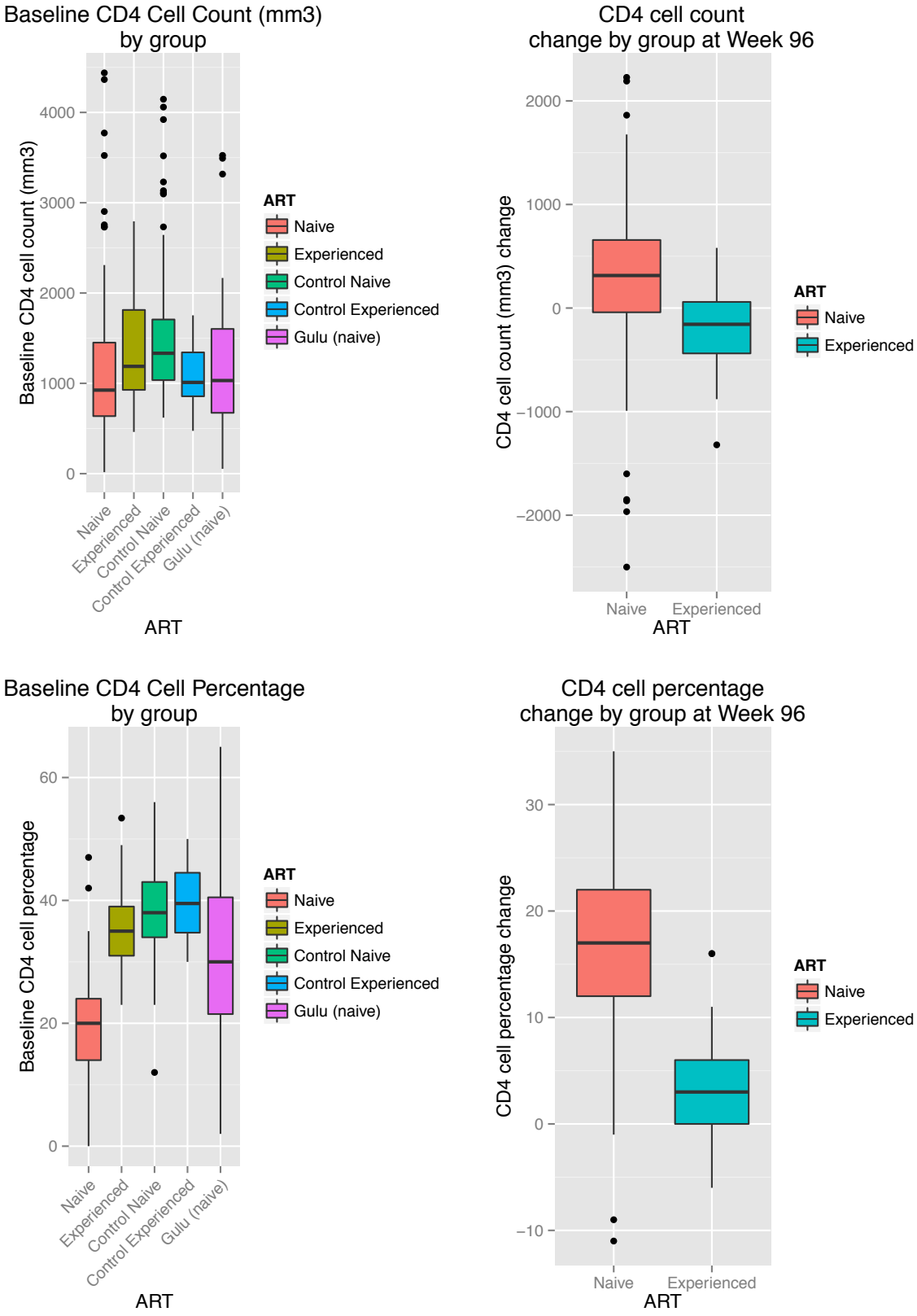


Figure 20 Boxplots of CD4 T cell counts and percentages over time by ART group

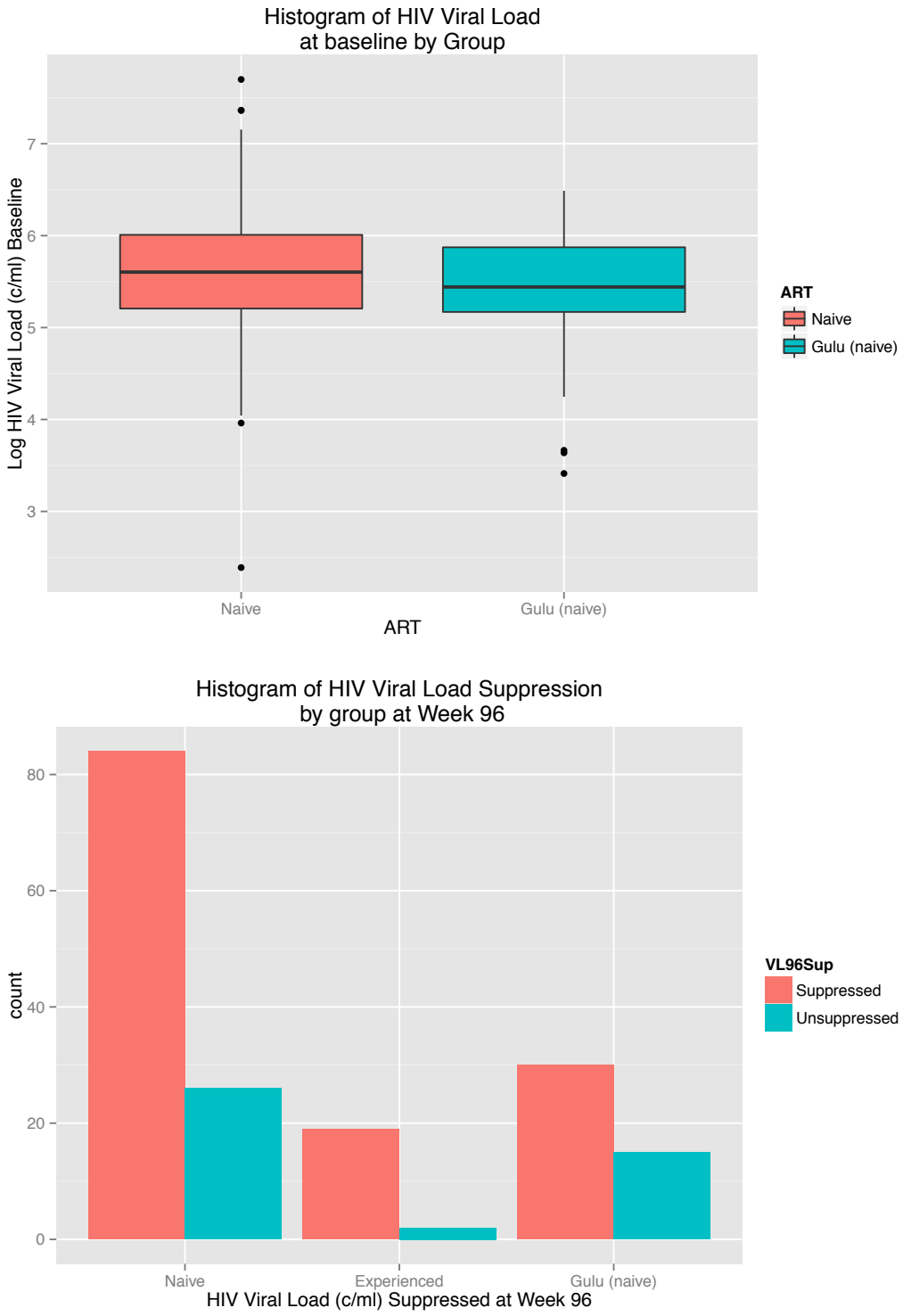


Figure 21 Histogram of viral load at baseline and suppression (<100 c/ml) by ART group at week 96

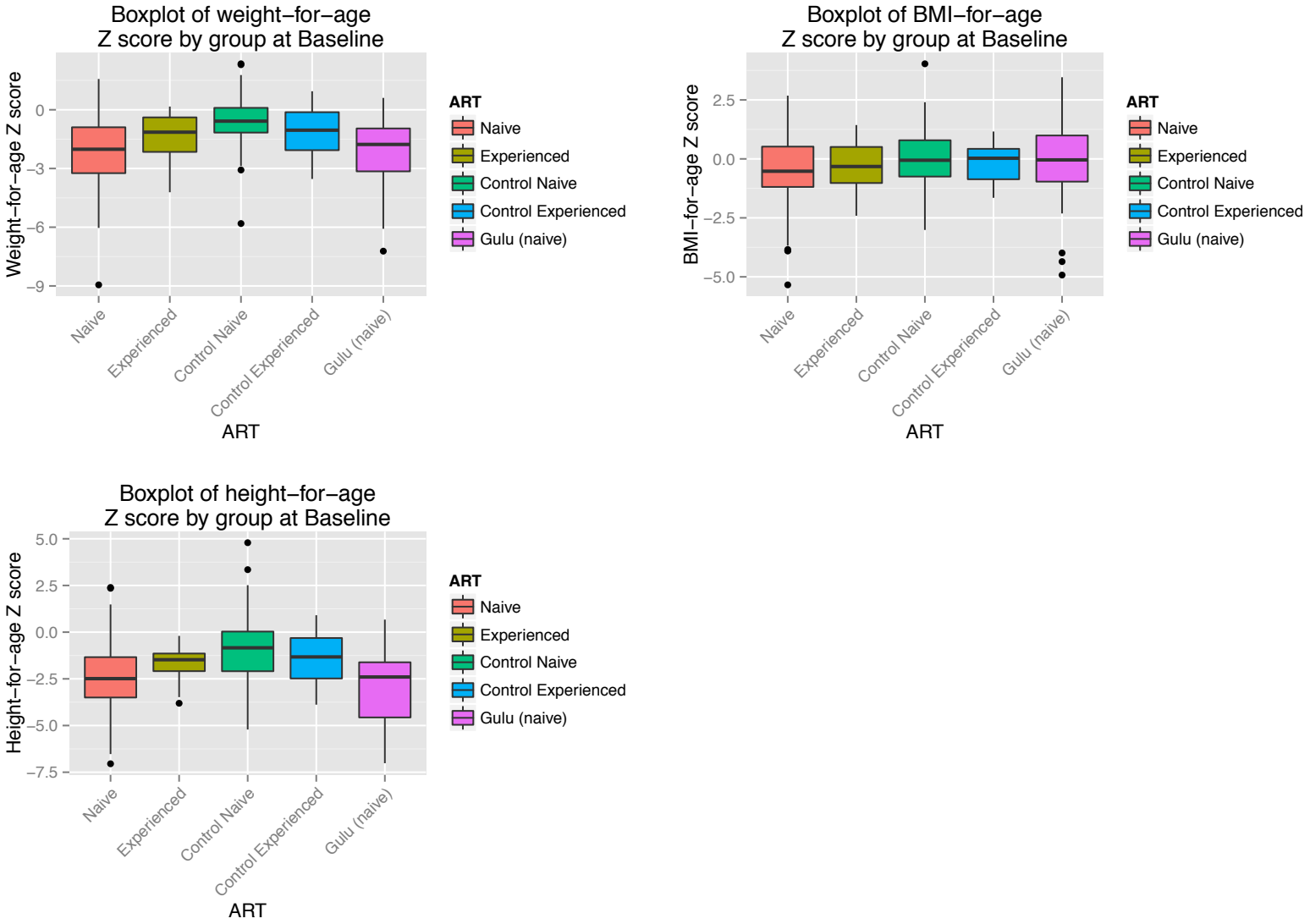


Figure 22 Boxplots of anthropometric markers at baseline by ART group

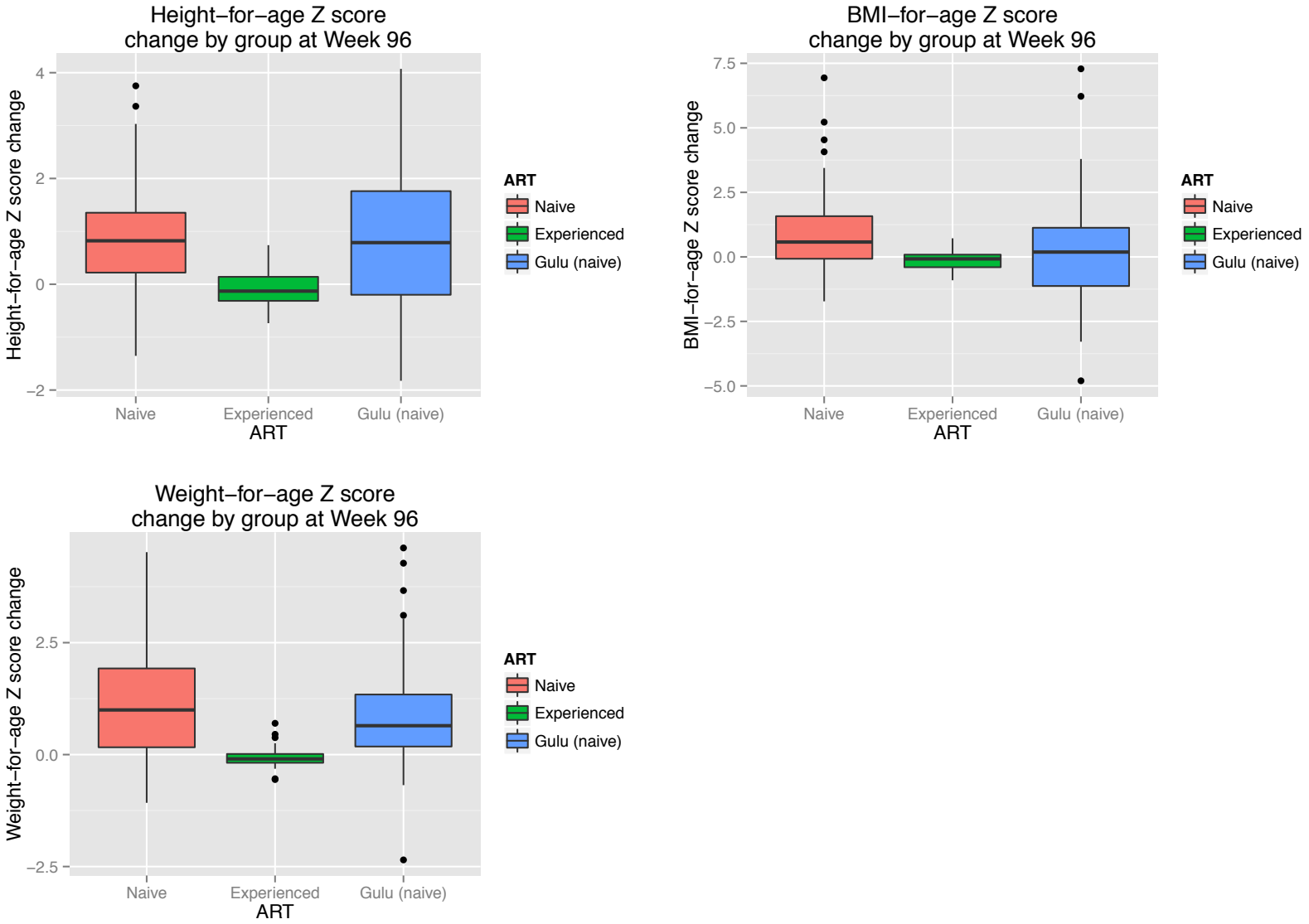


Figure 23 Boxplots of changes in anthropometric markers by ART group at week 96

Table 17 Age, ART randomisation and causes of death for 6 children who died

ART	Age at enrolment (years)	Duration of follow up (weeks)	Cause of Death
ART-naïve	1.3	38	Measles, pneumonia
Gulu (ART-naïve)	4.4	5	Pneumonia, malnutrition, hyperglycaemia, Gram negative sepsis
ART-naïve	3.1	25	Measles
Gulu (ART-naïve)	1.33	2	Unknown, died at home, probable pneumonia
Gulu (ART-naïve)	1.74	94	Unknown
ART-naïve	5	56	Kaposi's sarcoma

11.5 Discussion

This chapter describes the demographic and clinical features at enrolment of the children in the study cohort, and progress over time in the HIV-infected groups. It is reassuring to note that very few children were lost to follow up and that all HIV-infected groups had a high overall rate of viral suppression at week 96 (Figure 21). The number of deaths over 96 weeks for the whole group was low, 3/119 (3%) from the ART-naïve urban group and 3/55 (5%) from the ART-naïve rural Gulu group (Table 17). Overall within CHAPAS-3, there was minimal difference in outcome between the trial arms, and so differing impact of ART randomisation has not been further explored here.[45]

As viral load suppression was a precondition of enrolment in CHAPAS-3 for ART-experienced children, it is unsurprising that CD4 T cell count/percentage and anthropometric markers were higher at baseline for this group than the ART-naïve group. The ART-experienced group had already received a median of 4 years of ART (IQR 2.6-4.3) during which immune recovery and growth catch-up occurred. It is also unsurprising that the ART-naïve group made bigger gains in CD4 T-cell count/percentage and anthropometric markers over the course of the trial follow up.

The Gulu group were slightly younger than the ART-naïve group, had higher higher CD4 T cell percentage (Gulu group: median 30%, IQR 22-41, JCRC ART naïve group: median 20%, IQR 14-24; p<0.001) and slightly lower viral load (Gulu group: median 264700, IQR 130100-746300, JCRC ART naïve group: median 401200, IQR 161300-1021000; p=0.047). These findings could indicate that the Gulu children were brought for medical attention slightly earlier in their disease course than their urban peers. Anthropometric markers were similar in both ART-naive groups.

Considering the anthropometric markers in more detail, the only group with a weight-for age or height-for-age z score of 0 fell within the IQR was the HIV-negative, ART-naïve control group. The other, older control group had median weight-for age and height-for-age z scores of -1 (IQR -2.1- -0.1) and -1.3 (IQR -2.5- -0.3), and all HIV-infected groups had median z scores <0 for both weight and height-for-age. This indicates a base level of stunting within the study population compared with the reference population, but the reference population for anthropometric results was UK based which is a limitation of the study[46]. The WHO reference population, which itself is acknowledged to be flawed and is under review, did not include weights for the whole age range of the CHAPAS-3 children and so could not be used[47]. The gains seen in the ART-naïve groups' anthropometric markers are similar to those gains seen in other settings when initiating ART[48-50].

In general, all three groups of children with HIV did well over the course of the study with good CD4 T cell recovery and relatively high rates of viral suppression at 96 weeks. This is in line with outcome data from previous trials in similar settings[51-53]

12 Chapter 12 Quantitative PCR for the Detection of Specific Bacterial Species and Families as Indicators of Microbial Translocation

12.1 Background

As per section 8.4.6.3, qPCRs have the advantage of being less vulnerable to contamination than broad rand PCRs. This means that more cycles can be run rendering them more sensitive. However, they must be targeted to known bacteria and so may miss unexpected pathogens. A panel of assays was used targeted to bacteria implicated in microbial translocation.

12.2 Study Population

See section 9.1

12.3 Methods

Assays were run as per section 9.4.5 using appropriate positive and negative controls depending on primer set used. Negative to strong positives were calculated by comparison with standards of known CFUs and coded into discrete groups from 0-5 as per Table 10 reproduced here for reference. Proportion of positive and negative samples and the proportions of samples falling in each category from negative to strong positives (coded from 0-5) were compared using chi-squared tests for each assay between ART groups.

Table 10 Scheme for interpretation and comparison of PCR results

Normal Cut off PCRs	Low Cut off PCRs (Ent/TUF)	CFU equivalent	Ranking	Coded
>40.1	>35.1	0	Negative	0
37.0-40	32.0-35	1-10	Very low	1
34.0-36.9	29-31.9	10.1-100	low	2
31.0-33.9	26-28.9	100.1-1000	moderate	3
28.0-30.9	22.0-25.9	1000.1-10000	high	4
<28	<22	>10000.1	Very high	5

12.4 Results

Results were analysed and compared between assays using Table 10. In total, 668 plasma samples were available as per Table 18. 132 plasma and pellet samples were matched, with 8 unmatched pellet samples and no unmatched plasma samples.

Table 18 Sample availability by group at different time points

	Baseline	Week 12	Week 24	Week 72
ART-naïve (plasma)	112	108	7	112
ART-experienced (plasma)	22	19	0	22
ART- naïve (pellet)	118			
ART-experienced (pellet)	22			
Gulu *	52	49	3	49
Controls (Age matched ART-naïve)*	93			
Controls (Age matched ART-experienced*	20			

*Plasma samples only available.

In terms of missing samples, the ART-naïve group had 9/119 (8%) samples missing (not available) at baseline, and in the Gulu (ART-naïve) group 4/55 (7%) were missing. The control groups and the ART-experienced group were missing no samples

(Appendix D Missing samples across groups and time points). At week 12, the proportion of missing samples were 11/119 (9%) (ART-naïve), 3/22 (14%) (ART-experienced), and 5/53 (9%)(Gulu). At Week 72, excluding 5 children who had died and 4 who were lost to follow up, proportions of missing samples were 2/112 (2%)(ART naïve), 0/22 (ART experienced) and 7/51 (14%)(Gulu). The children who had missing samples were similar in age and baseline CD4 T cell counts to the other children in their groups. There were no assay failures in this group.

The majority of plasma samples were negative at all time points for most assays (Figure 24 - Figure 27, Table 19 - Table 21). The exceptions were *Enterobacteriaceae*, *S. aureus* and *Staphylococcus spp* although the levels seen tended to be very low even in these assays. Very few samples recorded strong positive results (higher CFU equivalents or code 4/5 as per Table 10) at any time point (Figure 24 - Figure 27). Samples positive for *Bifidobacterium, Lactobacillus* and Fusobacterium were seen but were extremely rare. At baseline, over half the Gulu samples were positive for *S. aureus* (55%, 95% CI 40-69%)(Table 19, Table 22, Figure 28), but this high proportion was not replicated in the other groups, and indeed was significantly higher than the other two HIV-infected groups (Table 19).

Approximately one-third of the ART-experienced group (38%, 95% CI 21-48%) and the Gulu group (33%, 95% CI 21-48) had positive results for *Enterobacteriaceae,* with 23% (95% CI 16-32%) of the ART-naïve group also having a positive result (Table 23). However those that were positive tended to be in the lower range at all time points and across all groups (Figure 24-Figure 27). The Gulu samples were also significantly more likely to be positive in more than one assay than the other two groups ($p<0.0001$), with 13.5% positive for *S. aureus* and *Staphylococcus spp.* and 9.6% for both *S.aureus* and *Enterobacteriaceae.*

In comparison with of the ART-naïve and experienced groups with age-matched HIV-uninfected control groups, the naïve-control group had a higher proportion of samples positive for *Enterobacteriaceae* than the HIV-infected ART-naïve group (37%, 95% CI 27-48% versus 23%, 95% CI 16-32, $p=0.03$)(Table 19)(Figure 29). The

experienced control group also had a higher proportion of positive samples but this did not reach significance (55%, 95% CI 32-77% versus 38%, 95% CI 18-62%, p=0.09). There were no other significant differences between the case and control groups in either proportions of positive versus negative samples or in proportions coded from negative to strong positive, although there was a trend towards the ART-naïve having a higher proportion of strong positives for *S.aureus* (p=0.06) than age-matched HIV-uninfected controls.

12.4.1 Plasma Samples over Time

Figure 28 and Figure 29 summarise changes over time in proportions positive of the *S. aureus* and *Enterobacteriaceae* PCRs within the HIV infected groups.

At week 12, the proportions positive for *S. aureus* trended towards an increase in ART naïve children from 9% (95% CI 5-17%) to 17% (95% CI 11-26%), but had fallen from 23% (95% CI 8-45%) to 16% (95% CI 3-40%) (ART-experienced) and in the Gulu group had fallen significantly from 55% (95% CI 40-69%) to 16% (95% CI 7-31%)(Table 20, Figure 28).

Proportions positive for *Enterobacteriaceae* had also increased in the ART naïve group from 23% (95% CI 15-32%) to 36% (95% CI 27-46%)(p=0.06) and decreased in both ART-experienced (from 38% (95% CI 18-62%) to 26% (95% CI 9-51%), p=0.64) and in the Gulu group from 33% (95% CI 21-48%) to 18% (95% CI 9-31%)(p=0.12)(Table 23)(Figure 29). There was a trend towards the ART naïve group having a higher proportion of samples positive for *Enterobacteriaceae* than the other two groups at this timepoint but this did not reach statistical significance (p=0.07). There were no other significant differences between groups at this point, including in comparing negative to strong positives between groups.

Very few samples were available at week 24, and only seven in the ART-naïve and three in the Gulu group. Only one sample was positive for *S.aureus* in each group and one for *Enterobacteriaceae* in the ART-naïve group.

At week 72, *S. aureus* results tended to be lower in all three groups than at week 12 and baseline: 6% (95% CI 3-12%) in ART-naïve (p=0.44 in comparison to baseline), 5%

(95% CI 0-23%)(p=0.188) in ART-experienced and 7% (95% CI 1-19%)(p<0.0001) in the Gulu group (Table 21, Table 22). *Enterobacteriaceae* were also present in 31% (95% CI 23-41%) of the ART-naïve, 23% (95% CI 8-45%) of the ART-experienced and 32% (95% CI 19-48%) of the Gulu group (Table 23). Again, there were no significant differences in the negative to strong positives between groups.

In general, with the exception of *S.aureus* at baseline in the Gulu group, proportions positive of both *S.aureus* and *Enterobacteriaceae* tended to fluctuate non-significantly over time and between groups (Figure 28, Figure 29)

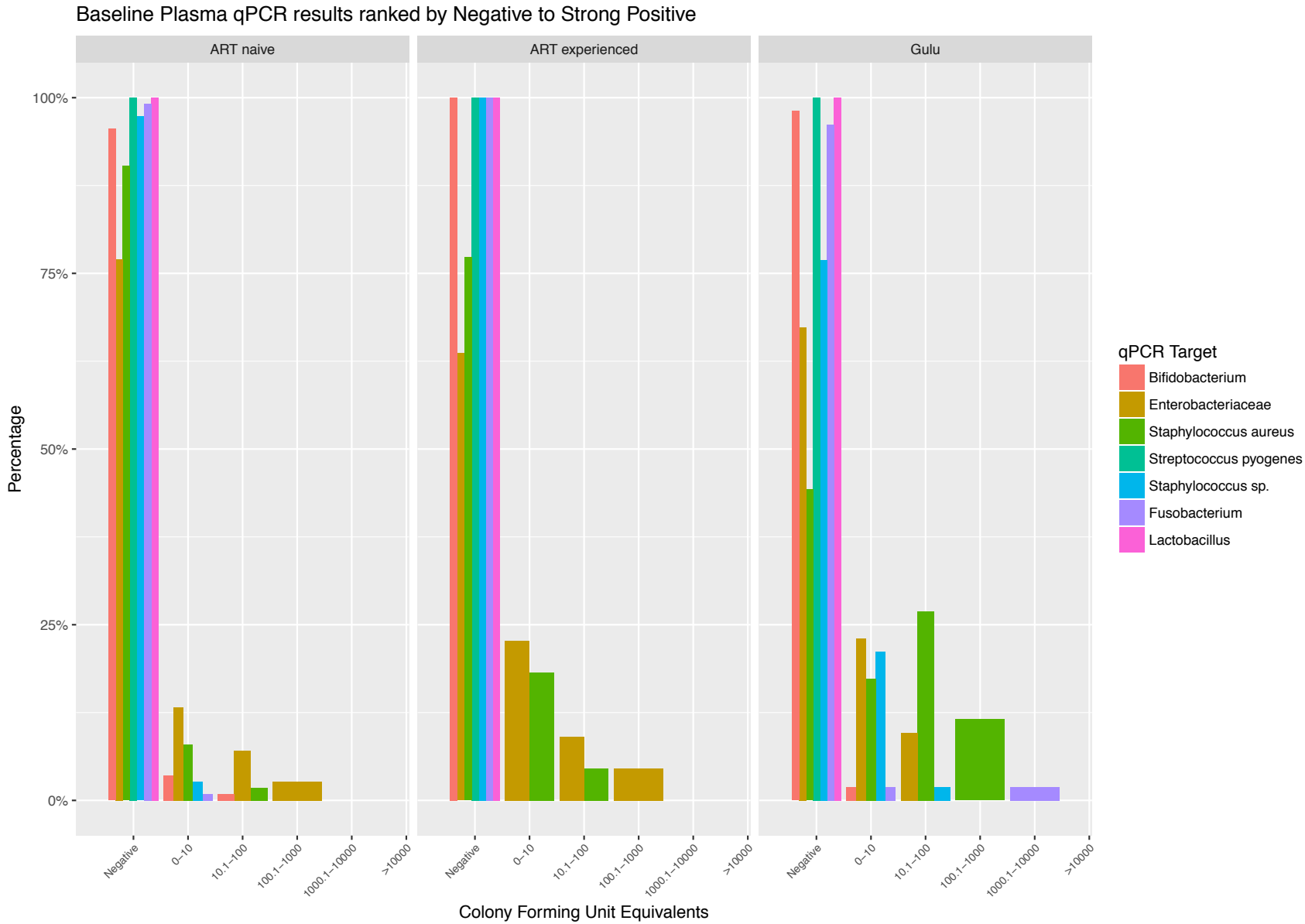


Figure 24 Baseline qPCR results ranked negative to strong positive by ART group

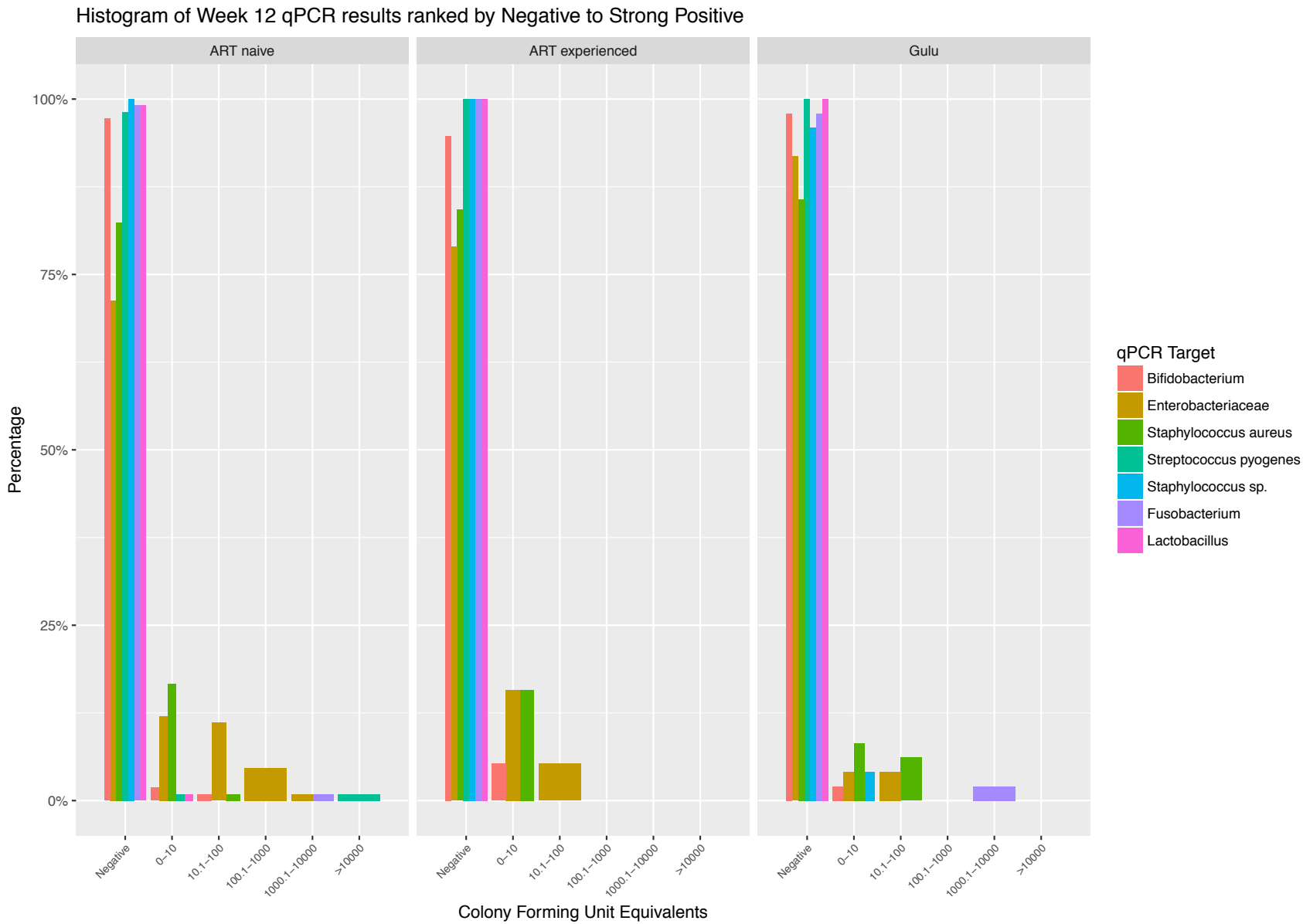


Figure 25 Week 12 qPCR results ranked negative to strong positive by ART group

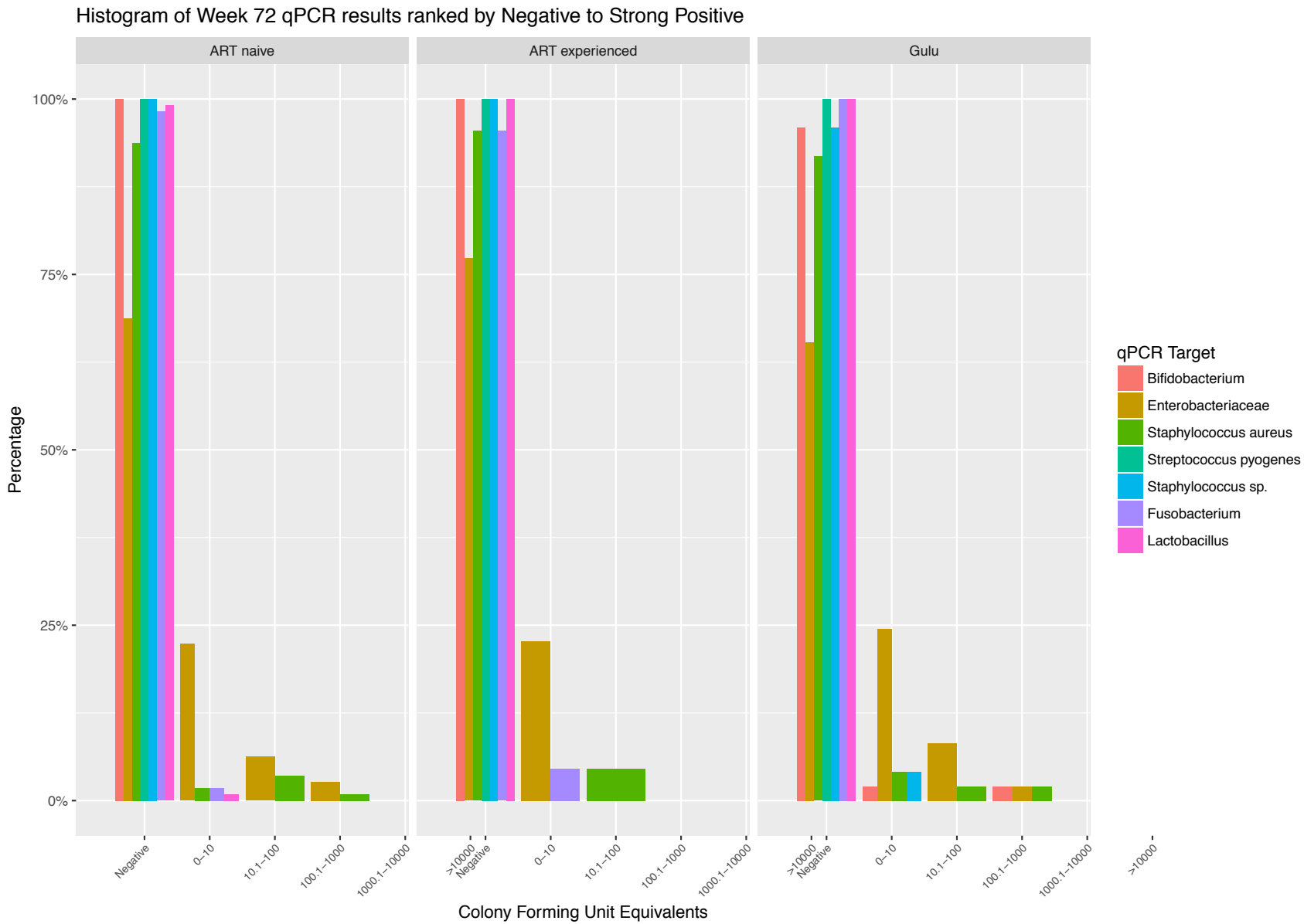


Figure 26 Week 72 qPCR results ranked negative to strong positive by ART group

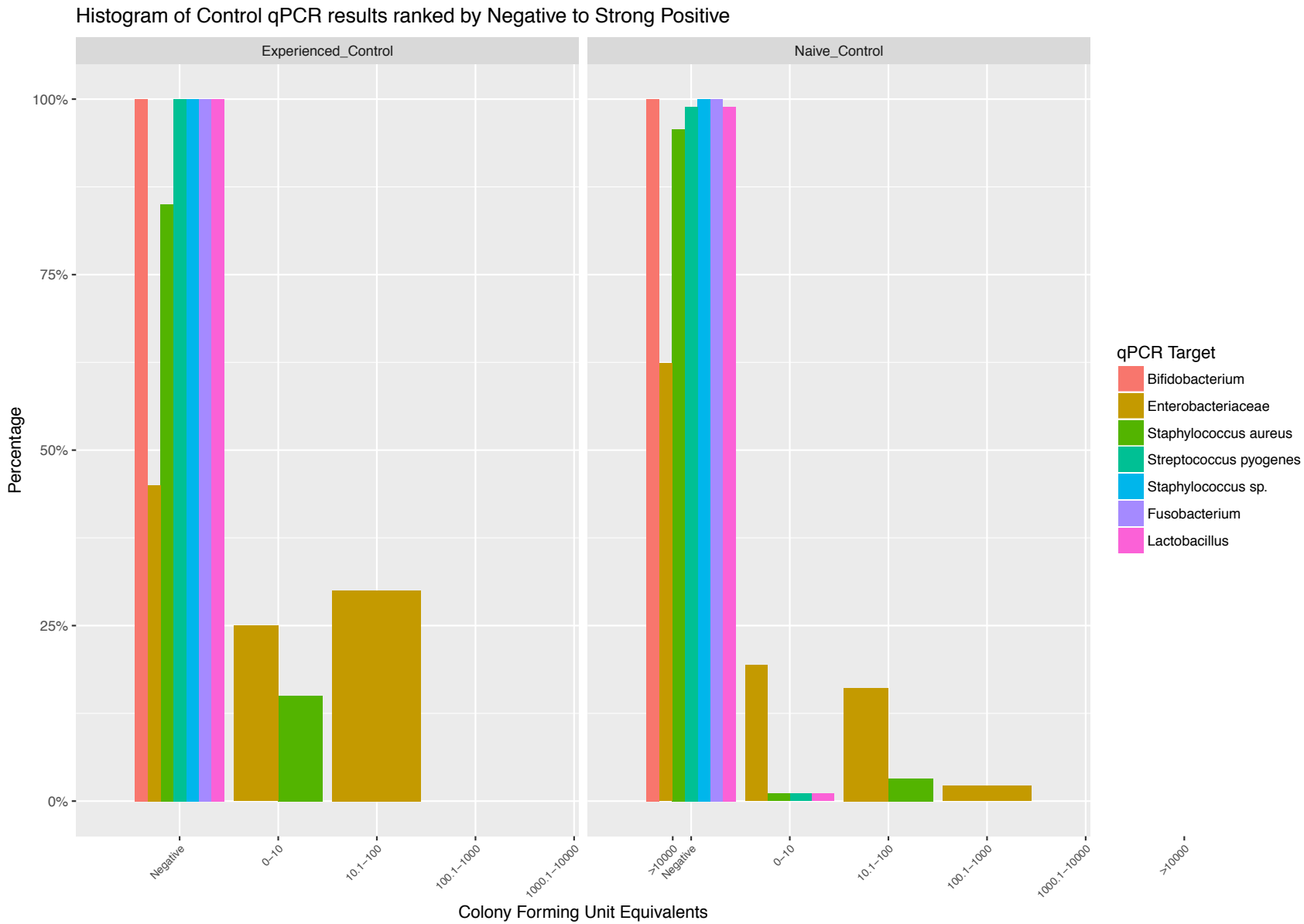


Figure 27 Control qPCR results ranked negative to strong positive by ART group

Table 19 qPCR results for samples at baseline comparing proportions of positive samples by ART group

	Positive Naïve N (%) n=112	% Positive Experienced N (%) n=22	% Gulu (Naïve) N (%) n=52	P Value (Fisher's exact, comparison across 3 ART groups)	% Control (naïve) N (%) n=93	% Control (experienced) N (%) n=20	P Value (Fisher's exact) proportion positive /negative cases versus matched controls	
							Naïve	experienced
Bifidobacteria	5 (4)	0	1 (2)	0.71	0	0	0.07	1
S. aureus	11(9)	5 (23)	29 (55)	<0.0001	4 (4)	3 (15)	0.42	0.61
S. pyogenes	0	0	0	1	1 (1)	0	0.45	1
Fusobacterium	1 (1)	0	2 (4)	0.47	0	0	1	1
Enterobacteriaceae	26 (23)	8 (38)	17 (33)	0.23	35 (37)	11 (55)	**0.03**	0.35
Staphylococci	3 (3)	0	12 (23)	<0.0001	0	0	0.25	1
Lactobacillus	0	0	0	1	1 (1)	0	0.45	1
Any mixed				<0.0001			1	0.49
S. aureus, Staphylococci, & *Fusobacterium*	1 (1)							
S. aureus, Staphylococci & *Enterobacteriaceae*			1 (2)					
S. aureus, Fusobacterium, & *Enterobacteriaceae*			1 (2)					
S. aureus, Staphylococci			7 (14)					
S. aureus Enterobacteriaceae	2 (2)	2 (9)	5 (10)		2 (2)			
Bifidobacterium, Staphylococci	1 (1)							
S. aureus,& *Fusobacterium*			1 (2)					
Staphylococci, Enterobacteriaceae			2 (4)					
Bifidobacteria, Enterobacteriacea			1 (2)					
Enterobacteriaceae, S.pyogenes					1 (1)			

Table 20 Week 12 qPCR results comparing proportions of positive samples between ART groups.

	% Positive Naïve (Week 12) N (%) n=108	% Positive Experienced (Week 12) N (%) n=19	% Gulu (Naïve) (Week 12) N (%) n=49	P Value (Fisher's exact) proportion positive/ negative across 3 ART groups
Bifidobacteria	3 (3)	1 (5)	1 (2)	0.62
S. aureus	19 (18)	3 (16)	8 (16)	0.95
S. pyogenes	2 (2)	0	0	1
Fusobacterium	1 (1)	0	0	0.62
Enterobacteriaceae	39 (36)	5 (26)	9 (18)	0.07
Staphylococci	0	0	2 (4)	0.15
Lactobacillus	0	0	0	1
Any Mixed				1
Bifidobacteria, S. pyogenes, Fusobacterium, & *Enterobacteriaceae*	1 (1)			
S. aureus, Staphylococci			1 (2)	
S. aureus Enterobacteriaceae	8 (7)	1 (5)	2 (4)	

Table 21 Proportions of positive qPCR results at week 72 by ART group

	% Positive Naïve (Week 72) N (%) n=112	% Positive Experienced (Week 72) N (%) n=22	% Gulu (Naïve) (Week 72) N (%) n=49	P Value (Fisher's exact) proportion positive across 3 ART groups
Bifidobacteria	0	0	2 (4)	0.15
S. aureus	7 (6)	1 (5)	3 (7)	1
S. pyogenes	0	0	0	1
Fusobacterium	2 (2)	1 (5)	0	0.47
Enterobacteriaceae	35 (31)	5 (23)	16 (32)	0.65
Staphylococci	0	0	2 (4)	0.15
Lactobacillus	0	0	0	1
Any Mixed				0.81
S. aureus, Staphylococci & Enterobacteriaceae			1 (2)	
S. aureus Enterobacteriaceae	1 (1)			
Bifidobacteria, Enterobacteriacea			1 (2)	

Table 22 Change in proportions positive for *S.aureus* over time by ART group

	Baseline (%, 95% CI)	W12 (%, 95% CI)	Fisher's exact test of proportions positive (Baseline versus W12)	W72 (%, 95% CI)	Fisher's exact test of proportions positive (Baseline versus W72)
ART-naive	9 5-17	18 11-26	0.118	6 3-12	0.336
ART-experienced	23 8-45	16 3-40	0.70	5 0-23	0.185
Gulu (ART-naïve)	55 40-69	16 7-31	<0.0001	7 1-19	<0.0001
Matched naïve control	4 1-11				
Matched experienced control	15 3-38				

Table 23 Table comparing proportions positive for *Enterobacteriaceae* over time by ART group

	Baseline (%, 95% CI)	W12 (%, 95% CI)	Fisher's exact test of proportions positive (Baseline versus W12)	W72 (%, 95% CI)	Fisher's exact test of proportions positive (Baseline versus W72)
ART-naïve	23 16-32	36 27-46	0.06	31 23-40	0.26
ART-experienced	38 18-62	26 9-51	0.64	23 8-45	0.44
Gulu (ART-naïve)	33 21-48	18 9-31	0.12	32 19-48	1
Matched naïve control	37 27-48				
Matched experienced control	55 32-77				

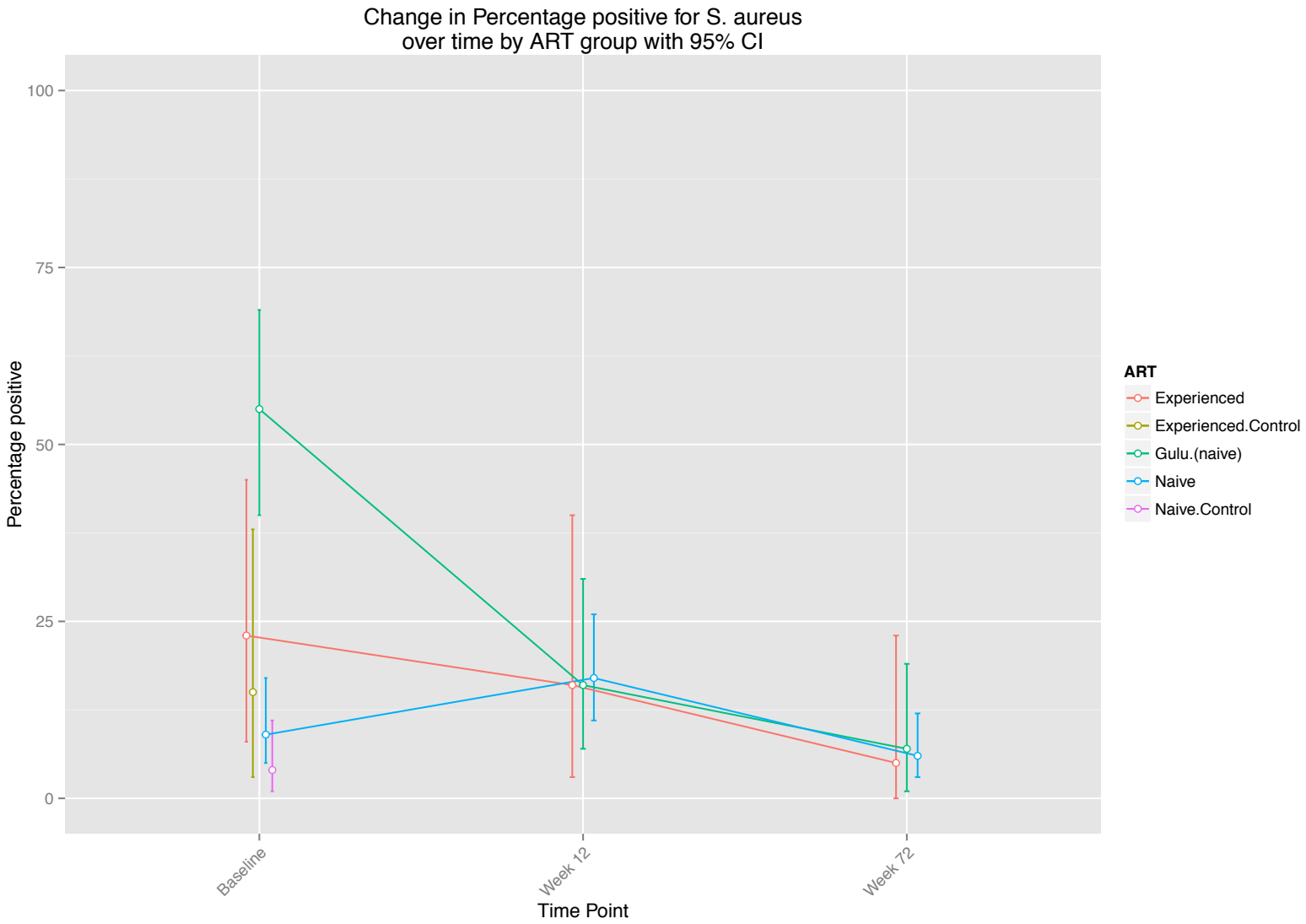


Figure 28 Change in proportions positive for *S.aureus* using qPCR assay over time with 95% CIs by ART group

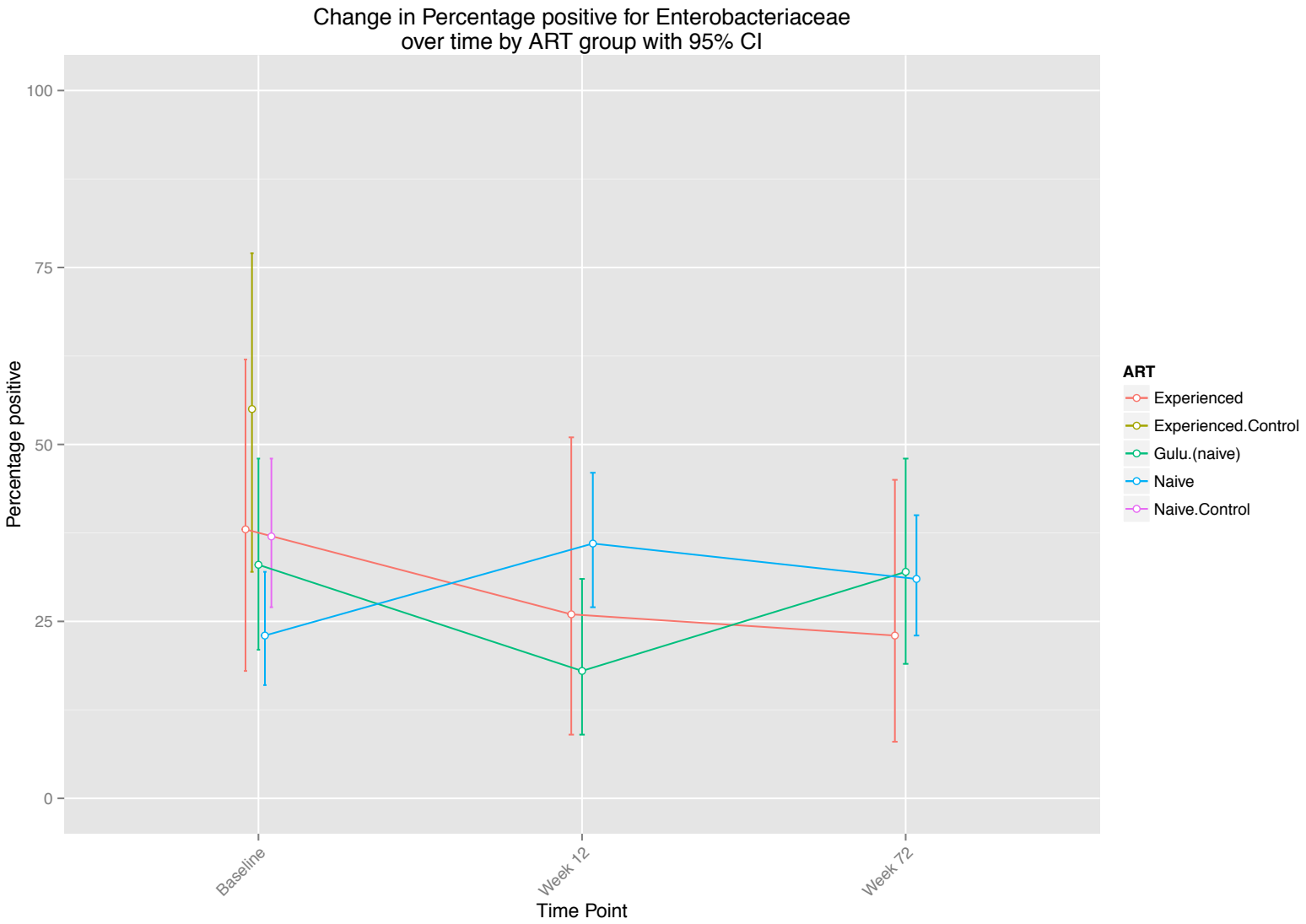


Figure 29 Change in proportions positive for *Enterobacteriaceae* using qPCR assay over time with 95% CIs by ART group

12.4.2 Pellet Samples

There were 140 pellet samples for analysis, 118 from ART-naïve children at baseline, and 22 from ART-experienced children (Table 24, Figure 30). The proportions of samples positive for *S. aureus* and *Enterobacteriaceae* were higher in both groups than in the plasma samples, and there were more samples with a higher quantity of bacteria recorded although levels still tended to be low (Figure 30). For ART naïve 52% (95% CI 42-61%) were positive for *S.aureus* versus 10% (95% CI 5-17%) in plasma, and for *Enterobactericeae* 69% (95% CI 59-77%) were positive in pellets versus 23% (95% CI 15-32%) in plasma. For the ART experienced group, 64% (95% CI 41-83%) were positive for *S.aureus* versus 23% (95% CI 8-45%) in plasma samples, and 59% (95% CI 36-79%) for *Enterobacteriaceae* versus 36% (95% CI 18-62%) in plasma, showing a trend towards higher proportions although not reaching statistical significance.

There was no significant difference between either proportions of samples positive between the two groups or proportions of samples testing weakly or strongly positive as per Table 10 (Table 24).

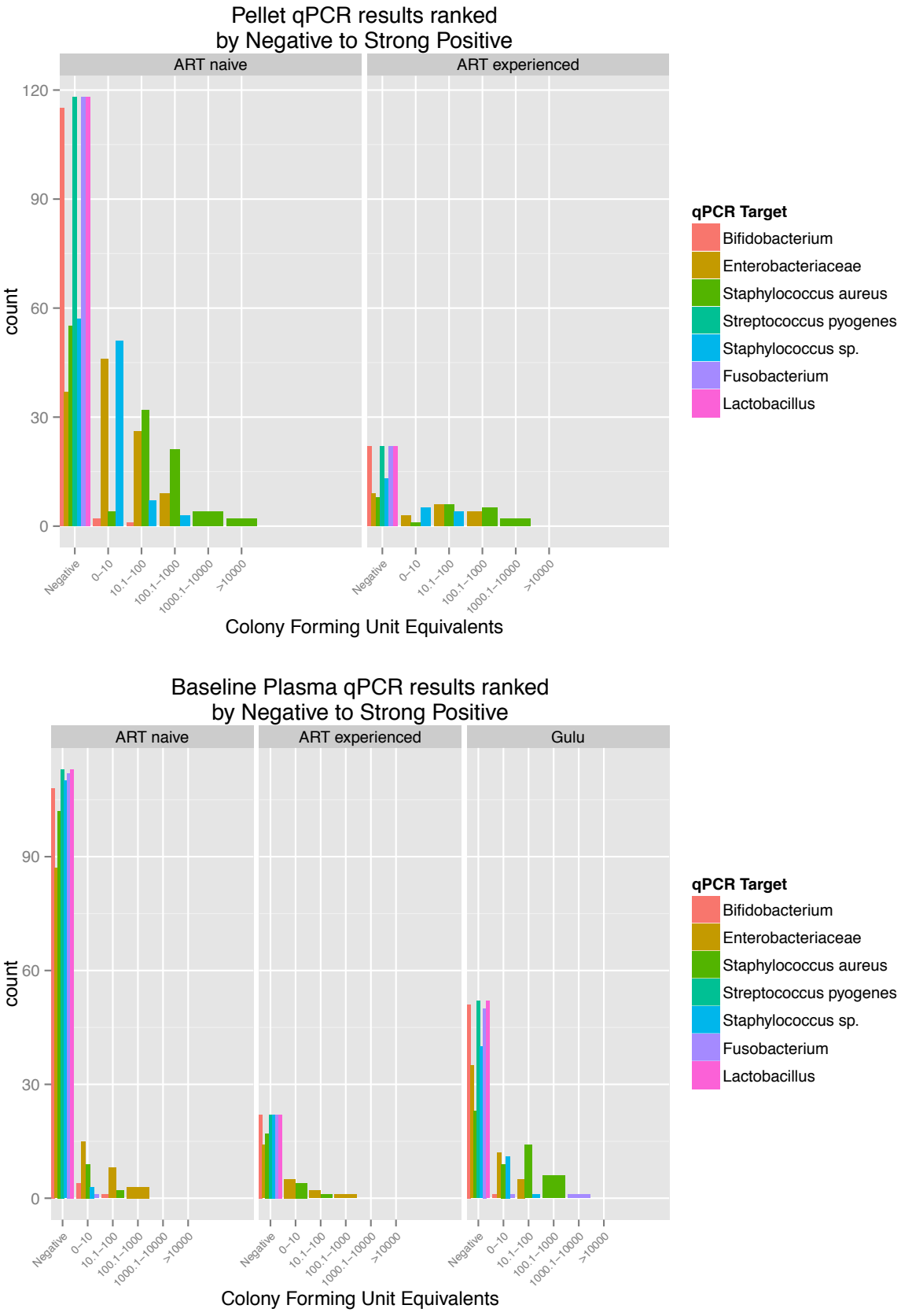


Figure 30 qPCR results at baseline comparing pellet with plasma samples by ART group: CFU equivalents as compared with CT values of standards with known CFU quantity and categorised from negative to strong positive.

Table 24 Table comparing qPCR results for cell pellets in HIV-infected children at baseline by ART group

	Positive Naïve (NA=1) N (%) n=118	% Positive Experienced N (%) n=22	P value (2 sided Fisher's exact test) comparing positive versus negative	P value 2 sided Fisher's exact test proportions low- high CFUs
Bifidobacteria	3 (3)	0	1	1
S. aureus	61 (52)	14 (64)	0.36	0.66
S. pyogenes	0	0	1	1
Fusobacterium	0	0	1	1
Enterobacteriaceae	81 (69)	13 (59)	0.46	0.06
Staphylococci	61 (52)	9 (41)	0.49	0.09
Lactobacillus	0	0	1	1
Any mixed	72 (61)	11 (50)	0.35	
S. aureus, Staphylococci* & *Enterobacteriaceae	25 (21)	4 (18)		
S. aureus* & *Staphylococci	14 (12)	1 (5)		
S. aureus* & *Enterobacteriaceae	9 (8)	2 (9)		
Bifidobacteria, S. aureus,* & *Enterobacteriaceae	1 (1)			
Staphylococci, Enterobacteriaceae	22 (19)	4 (18)		
Bifidobacteria, Enterobacteriacea	1 (1)			

12.5 Discussion

To our knowledge, this is the first use of qPCR for specific bacterial species/families to investigate microbial translocation in the blood of children with HIV. Our aim was to add sensitivity (not achievable with broad range 16S rDNA PCR) for particular bacterial species/families such as *Enterobacteriaceae* previously implicated in microbial translocation in HIV infection[54,55], or known gut commensals such as *Bifidobacterium* and *Lactobacillus,* the levels of which in the gut microbiome might be affected by HIV infection.

However, most of the qPCR plasma assays were negative or very low at most time points, including *Bifidobacterium*, *Fusobacterium, Lactobactillus,* and *S. pyogenes. S. aureus* and *Enterobacteriaceae* were positive in more samples, with the highest proportion of positive plasma samples being recorded in the Gulu group at baseline (55% for *S.aureus*, 95% CI 40-69%) (Figure 24, Table 22). The baseline pellets had significantly higher proportions of positive results than plasma for both *S.aureus* and *Enterobacteriaceae* in the ART naïve group, with a trend towards higher proportions in the ART experienced group though confidence intervals were wide (Figure 30).

The *Staphylococcus spp*. assay covers the whole family and was more vulnerable to contamination that the specific *S.aureus* qPCR, as evidenced by negative extraction and negative control samples crossing the cycle threshold <40 cycles. Therefore the threshold for a positive result in this assay and the *Enterobacteriaceae* assay was set lower at 35 cycles rather than 40 cycles as for the other assays. The loss of sensitivity, meant that some samples that were positive for the *S. aureus* assay were not for the *Staphylococcus spp.* Similarly, samples positive for both *S.aureus* and *Staphylococcus spp.* are likely to represent the presence of *S. aureus* alone. This is seen most clearly in the results from the pellet samples where the proportion of ART naïve samples positive for both *S.aureus* and *Staphylococcus spp.* were the same (52%, 95% CI 42-61%), although in the ART experienced group the proportions positive for *Staphylococcus spp*. were lower than *S. aureus.* For reasons discussed below (Section

14.5), pellet samples appear to have been more likely to contain bacterial DNA than plasma samples, but it should be noted here that although the levels of DNA detected were higher in the pellet than plasma samples (possibly secondary to centrifugation), the species of bacteria identified in plasma and pellet samples were broadly similar. This runs counter to the *a priori* hypothesis that different bacterial species might be preferentially phagocytosed during the process of microbial translocation, although might indicate phagocytosis more generally of bacterial DNA in the blood.

In comparison with controls, it is interesting that there appeared to be significantly more naïve-control samples positive for *Enterobacteriaceae* than in the naïve group at baseline (37%, 95% CI 27-48%, versus 23%, 16-32% p=0.03), with a trend towards a similar pattern in the experienced versus experienced-control (Figure 30). There was a trend towards an increase in the proportion of positive samples in the urban ART-naïve group over time (31% at week 72, 95% CI 23-40%, p=0.26 in comparison with baseline) but this did not reach significance. Overall there was some fluctuation in the proportion of positive samples for *Enterobacteriaceae* in all three HIV-infected groups but no significant change over time (Figure 29). As can be seen in Figure 29 the 95% confidence intervals around the proportions were wide and overlapping between groups and over time. This could indicate either that the assay is working at the limits of detection and the very low results seen are in fact representative of contamination, or alternatively that there is some evidence of *Enterobacteriaceae* DNA present in both HIV-infected and uninfected individuals and also in those established and initiating ART. The latter interpretation could fit with the hypothesis that some baseline microbial translocation occurs ubiquitously, but as there was little difference between groups, it may be that host response is more important biologically than microbial translocation per se.

With regards to the *S. aureus* results, the Gulu (rural) group supported most closely the *a priori* hypothesis of microbial translocation being reduced by effective ART, with a significant reduction in the number of positive samples over time (baseline: 55%, 95% CI 40-69, Week 72: 7%, 95% CI 1-19%, p<0.0001). Results from the two JCRC

(urban) groups do not conform to this hypothesis, as results in the ART naïve group were lower than both other groups ($p<0.001$) at baseline, and although there was a non-significant increase at week 12, remained low at week 72 (6%, 95% CI 3-12%)(Figure 28). However, confidence intervals were wide, especially around proportions for the ART-experienced group due to small group size (Table 19)(Figure 28). In general, with the exception of the Gulu group, the proportions of samples positive for *S.aureus* appeared to fluctuate non-significantly over time with overlapping 95% confidence intervals as seen in Figure 28. As many of the positive results were low, these assays were working at the limits of detection, and so the results must be interpreted with caution. Furthermore, the *S.aureus* results from the rural Gulu population raise the question as to whether this is microbial translocation from the gut or from skin contamination that is being measured. *S.aureus* frequently colonises skin, and appears to be at a higher prevalence in rural than urban communities[56].

Data as regards previous use of qPCRs for specific bacterial species are extremely limited, so external comparisons are limited to the results of sequencing after broad range 16S rDNA PCR. Members of the *Enterobacteriaceae* family have previously been sequenced in this setting subsequent to broad range 16S rDNA PCR, and were identified in a small number of HIV-positive samples (6/12 ART-naïve adults with advanced HIV, 5/7 with poor response to ART, 0/7 in adults with a good response to ART[54]). However, no HIV uninfected controls were included, and the numbers were small. In a similar study by the same group which included 44 HIV infected patients and 13 HIV uninfected controls, cloning and Sanger sequencing after broad range 16S rDNA PCR revealed *Enterobacteriaceae* in up to 75% of samples, and a significant difference in *Lactobacillus sp.* and *Pseudomonas sp.* prevalence between those with a good immune response to ART and those without[8]. No samples were positive by broad range 16S rDNA PCR for the HIV-uninfected controls. No *Staphylococcus spp.* was identified in any group. However, no sensitivities are quoted for the assay, the broad range 16S rDNA PCR was run to 40 cycles, which would increase the risk of false positives due to contamination, and sequencing identification was limited to

family level. The group sizes were very small (between 4-8 at different time points) so the results must be interpreted with caution.

Some of these limitations also applied to this study, in that the ART-experienced group size was small, meaning many results have wide confidence intervals. There were some missing samples, although in relatively low proportions. The highest proportion of missing samples was in the Gulu group at week 72 (14%). However, in all three groups, the children for whom samples were missing were similar in characteristics (age and CD4 cell count/percentage) to those for whom samples were available. The higher proportion of missing samples from Gulu may be a question of logistics, in that samples had to be packaged for transport twice, once from Gulu to Kampala, then again from Kampala to London, meaning there were additional opportunities for samples to be lost. Alternatively, there may have been difficulties for the children in attending follow-up, so some visits may have been missed without the child being entirely lost to follow up for the study.

The qPCR assays are heterogenous with varying degrees of specificity and sensitivity, limiting direct comparison between assays. To overcome this and try to limit over interpretation I used a ranking scheme to compare results as per Table 10. The biggest limitation is that it appears that the pellet samples had a higher proportion of positive test results for both *S.aureus* and *Enterobacteriaceae,* but pellet samples were only available at baseline for the ART naïve and experienced groups from JCRC, so no comparison was possible with controls, or the Gulu group, or over time.

However, the overall pattern of results were similar between pellets and plasma, in that most assays (*Fusobacterium, S.pyogenes, Lactobacillus, Bifidobacterium*) were all negative, with *S.aureus* and *Enterobacteriaceae* being the assays that yielded the highest numbers of positive results throughout. *Enterobacteriaceae* in particular have been detected in the blood of HIV-infected individuals in previous studies, although they present challenges in terms of interpretation as the family contains species that are both potential pathogens such as *Escherichia coli*, and environmental species,

such as *Serratia spp*, that are rarely isolated from humans and are therefore more likely to be contaminants.[8,15,57,58] The family has undergone frequent reclassification as sequencing methods evolve, meaning even species identification can be challenging[57,58].

These results should be interpreted together with the results from NGS and broad range 16S rDNA PCR assays (see below). In summary however, despite using assays of high sensitivity minimal evidence was found of DNA from the *Bifidobacterium, Fusobacterium* , *S. pyogenes* or *Lactobacillus* assays in the blood of the study population. Although *Enterobacteriaceae* and *S.aureus* were identified, the levels detected were low in plasma samples at all time points in all groups, indicating these assays were working at the limit of detection (Figure 24 to Figure 27). Furthermore, for the *Enterobacteriaceae* assay, there was no consistent difference between groups or over time (Figure 29), so it is questionable as to whether these results have biological significance. The *S. aureus* did appear to decrease in the Gulu (ART-naïve) group over time (Figure 28), which could indicate a true change in translocating bacterial DNA from the gut. However, *S.aureus* is a common skin commensal, the assay is very sensitive, and levels detected were low. These results too should be interpreted with caution.

13 Chapter 13 Broad Range 16S rDNA PCR for the Detection of Bacterial DNA as an Indicator of Microbial Translocation in HIV-infected and Uninfected Children.

13.1 Background

As described in 8.4.6.1 and 8.4.6.2 a combination of broad range 16S rDNA PCR with NGS to identify the species present is a method to explore more comprehensively the bacterial DNA present in a sample, including rare or unexpected species, and groups of organisms that are difficult to target using a specific qPCR owing to high levels of genetic diversity. Furthermore 16S rDNA PCR is necessary to quantify the bacterial load in each sample prior to NGS in order to determine relative abundances in the NGS data generated. This chapter documents the results obtained using a broad range 16S rDNA PCR using a SYBR® Green dye.

13.2 Study Population

See section 9.1

13.3 Method

Samples were extracted as per section 9.3

Broad range 16S rDNA PCR was carried out using a SYBR® Green assay according to section 9.5. Analysis was carried out as per section 9.7. As a log scale was used for graphical representation, a constant of one was added to all results generated by the SYBR® Green broad range 16S rDNA PCR. When considering the broad range 16S rDNA PCR in isolation, CFU equivalents were used to estimate bacterial load which were calculated by comparison with standards of known CFUs. When comparisons were made with other assays such as the qPCRs the scheme described in Table 10 was used (see Section 10.1.5). Negative to strong positives were calculated by comparison with standards of known CFUs and coded into discrete groups from 0-5 as per Table 10. Proportion of positive and negative samples and the proportions of samples falling in each category from negative to strong positives (coded from 0-5) were compared using chi-squared tests between ART groups.

13.4 Results

Sample numbers available for analysis were the same as for section 12.4 (Table 18). Samples were missing in the same proportions as per the qPCR assays with the exception of one extra sample missing at week 72 in the ART-naïve group (3/112, 3%) (Section 12.4 and Appendix D Missing samples across groups and time points).

13.4.1 Broad Range 16S rDNA PCR using SYBR® Green Assay

For plasma samples, results were broadly similar across groups and time points(Figure 31). As seen in Figure 32 there was some fluctuation in levels, but IQRs of median CFU equivalents were broad and overlapping. Considering comparisons with controls, baseline results were very similar between both ART-naïve and age-matched controls (median CFU equivalents 103 (IQR 21-205) and 102 (IQR 45-148) respectively, p=0.47) and ART-experienced and controls (median CFU equivalents 133 (IQR 77-185) and 103 (IQR 53-152) respectively, p=0.25). The samples from Gulu had a significantly lower median (37, IQR 12-96) when compared with Wilcoxon rank sum tests to the naïve group at baseline (p=0.0015), but there were no other significant differences between the groups at any time point.

Considering changes in bacterial load of samples over time, paired Wilcoxon rank sum tests for each individual over time revealed some evidence of difference between the Gulu samples at baseline and week 12, and baseline and week 72 (Table 26). However, Spearman correlation coefficients (using complete observations) were small and not significant and there appeared to be minimal relationship between results in individuals over time (Table 27).

Comparing pellet with plasma samples, as with qPCR results, the bacterial loads in the pellet samples were significantly higher than for plasma overall ($p<0.001$ in paired Wilcoxon rank sum test in both groups) although there was no direct relationship between individual paired samples i.e. between the bacterial load detected in plasma and pellet samples from an individual (Spearman Correlation coefficient 0.07, p=0.44 in ART naïve, Spearman Correlation coefficient 0.003, p value=1 in ART experienced)(Figure 33). Bacterial loads detected in ART experienced and ART naïve

patients were similar as demonstrated by boxplot, also showing the difference between plasma and pellet results (Figure 33).

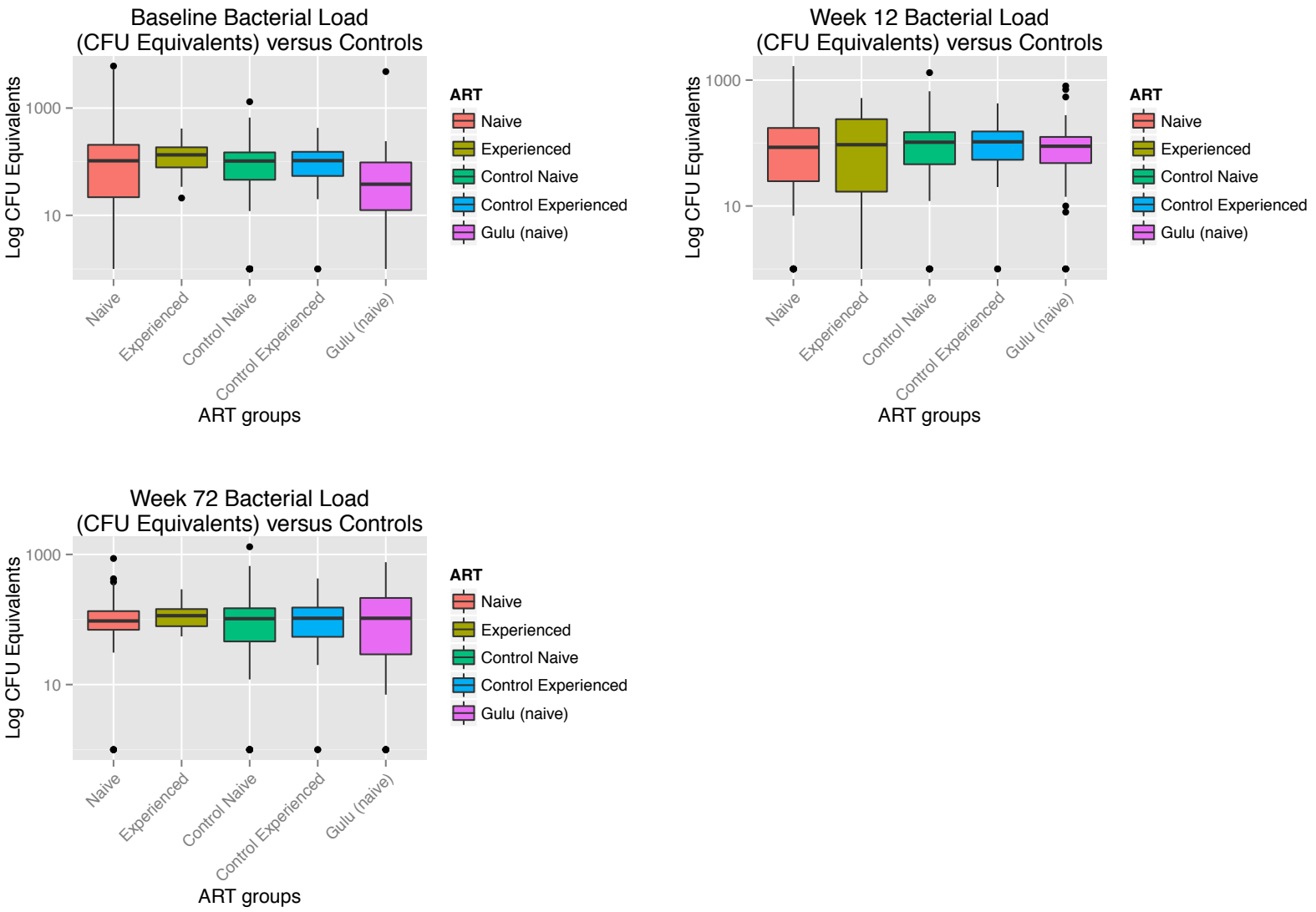


Figure 31 Boxplots of Bacterial load (CFU equivalents compared with standards of known CFUs) detected by broad range 16S rDNA PCR (SYBR® Green) by ART group over time

Table 25 Bacterial load (CFU Equivalents) results by broad range 16S rDNA PCR assay (SYBR® Green) by ART group over time compared using unpaired Wilcoxon Rank sum tests

	Naïve	Experienced	P value naïve versus experienced	Naïve Controls	Experienced Controls	P value cases versus controls	Gulu	P value (Gulu vs naïve cases)
Baseline CFU equivalent (Median, IQR)	103 21-205 n=112	133 77-185 n=22	0.37	102 45-148 n=93	103 53-152 n=20	0.47 (naïve) 0.25 (exp)	37 12-96 n=52	**0.0015**
W12 CFU equivalent (Median, IQR)	85 24-173 n=108	93 16-240 n=19	0.72	NA	NA	0.55 (naïve) 0.99 (exp)	88 47-124 n=49	0.94
W24 CFU equivalent (Median, IQR)	87 8-155 n=7	NA	NA	NA	NA	0.6 (naïve)	200 100-219 n=3	0.47
W72 CFU equivalent (Median, IQR)	94 69-134 n=112	114 78-144 n=22	0.36	NA	NA	0.56 (naïve) 0.62 (exp)	104 29-214 n=44	0.94

Table 26 Paired Wilcoxon Rank sum tests of bacterial load (CFU equivalents) over time by ART group

Paired Wilcoxon rank sum test		ART Naïve p value	ART Experienced p value	Gulu (naïve) p value
PreRandomisation	Week 12	0.072	0.8	**0.03**
PreRandomisation	Week 24	0.47	NA	0.5
PreRandomisation	Week 72	0.09	0.39	**0.02**
Week 12	Week 24	0.63	NA	1
Week 12	Week 72	0.28	0.98	0.19
Week 24	Week 72	0.25	NA	0.75

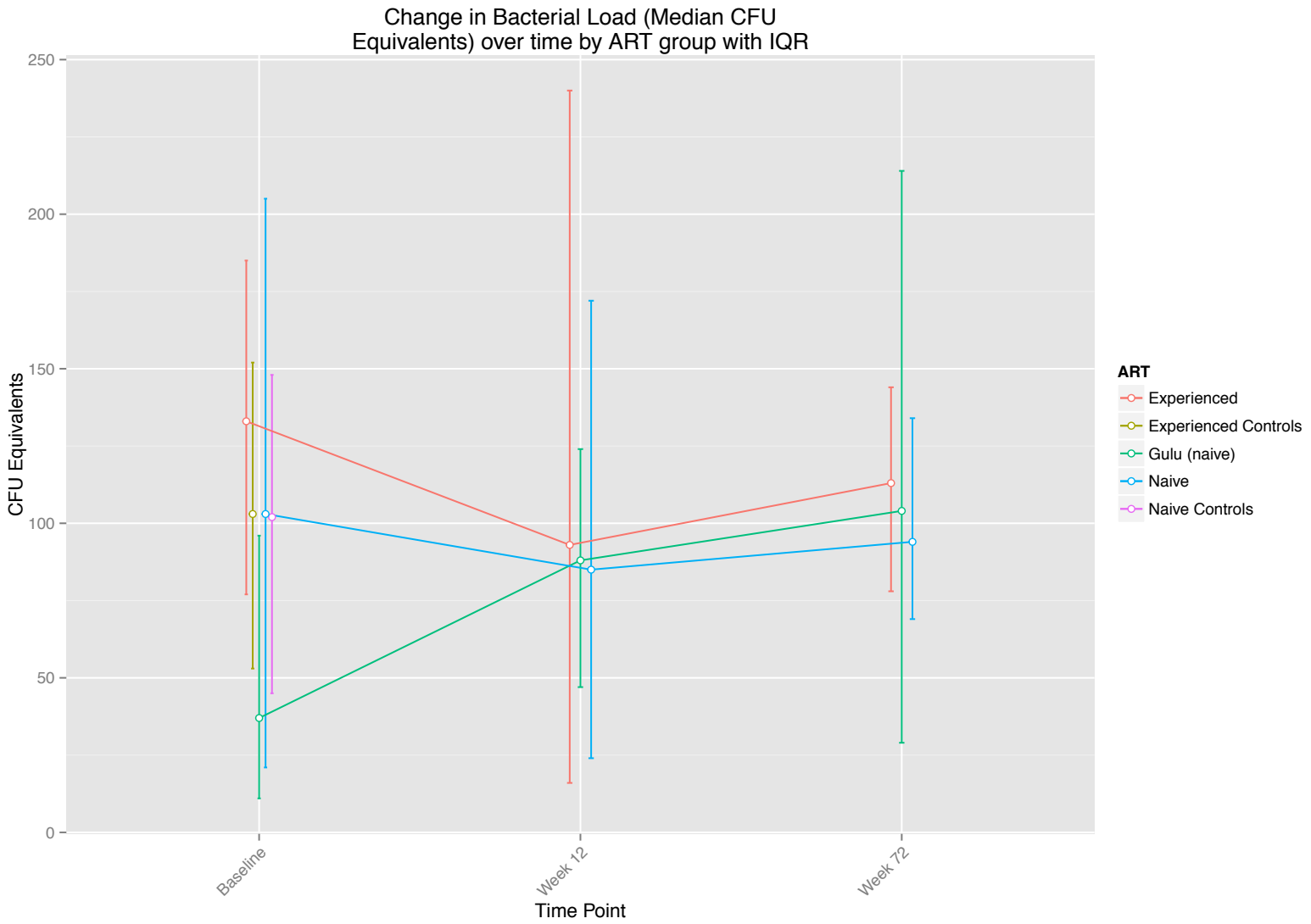


Figure 32 Change in bacterial load (median CFU equivalents compared with standards of known CFUs) by broad range 16S rDNA PCR over time by ART group with IQR

Table 27 Spearman Correlation Coefficients by ART groups over time

Spearman Correlation using complete observations		ART Naïve		ART Experienced		Gulu (naïve)	
		Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
PreRandomisation	Week 12	0.14	0.17	-0.08	0.75	-0.12	0.44
PreRandomisation	Week 24	-0.07	0.49	NA	NA	-0.87	0.33
PreRandomisation	Week 72	0.14	0.17	0.19	0.39	-0.04	0.81
Week 12	Week 72	-0.07	0.49	0.02	0.93	-0.08	0.63

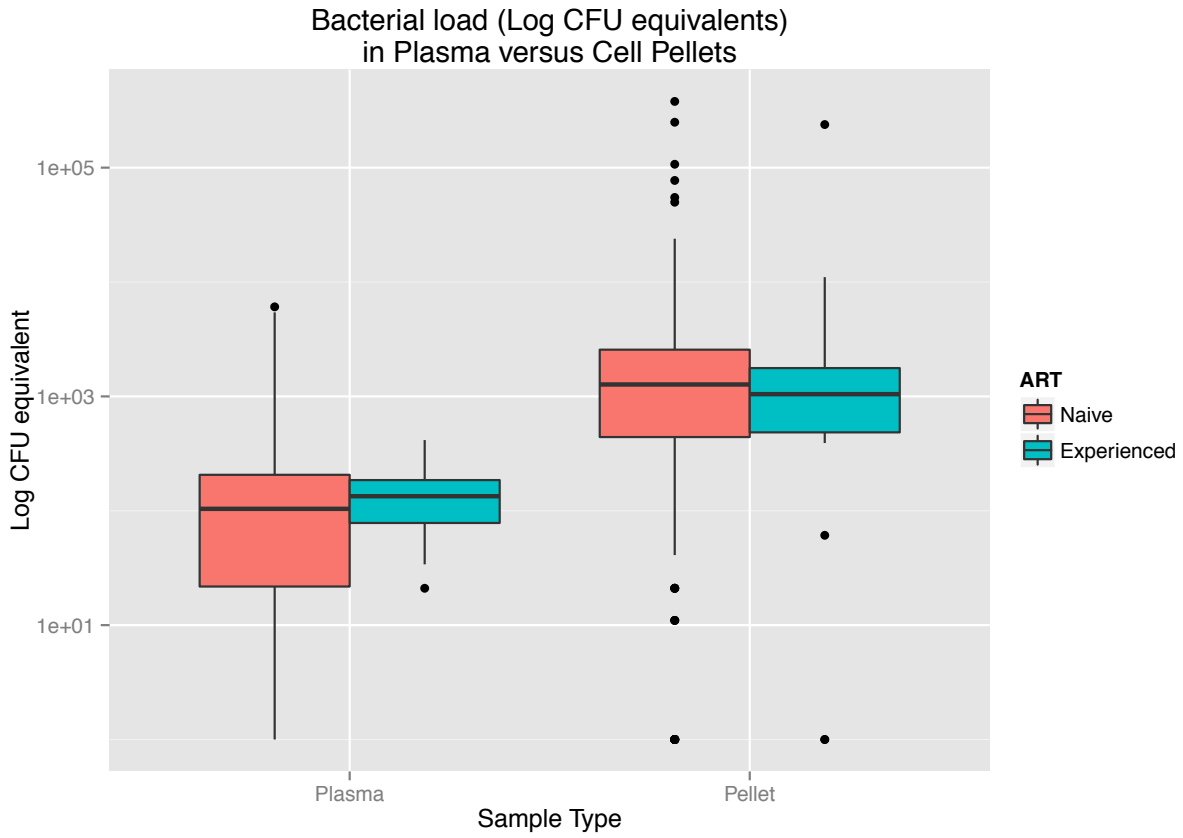


Figure 33 Comparisons of bacterial load (CFU equivalents compared with standards of known CFUs) in pellets versus plasma samples at baseline

13.4.2 Comparison of Broad Range 16S rDNA PCR (using SYBR® Green assay) with qPCRs

The results obtained from the broad range 16S rDNA PCR (SYBR® Green) were compared with those from the individual qPCRs using the scheme set out in Table 10 (see section 10.1.5), categorising results into discrete groups from negative to very high results as compared with standards of known CFUs. Very few samples tested positive for *Fusobacterium, Bifidobacterium, Lactobacillus*, or *S.pyogenes* and very little relationship was seen between the results of these assays and the broad range 16S rDNA PCR (SYBR® Green)(data not shown). With *S. aureus* and Staphylococcus spp. there was minimal relationship between assays, but with *Enterobacteriaceae* there appeared to be an inverse relationship between the results obtained by qPCR and broad range 16S rDNA PCR which was significant when comparing proportions of negative to positives, although the decrement was small ($p=0.03$ using Fisher's exact test) (Figure 34). ART group appeared to impact minimally on the comparison (data not shown).

As regards pellet samples, there was a significant positive relationship between samples positive for *S. aureus* and those with a positive result via broad range 16S rDNA PCR (SYBR® Green) (Fisher's exact test $p=0.023$)(Figure 35). However there was once more a significant negative relationship between assays positive for *Enterobacteriaceae* and using the broad range assay (Fisher's exact test $p=0.02$) (Figure 35).

To investigate a possible relationship between 16S broad range PCR results (SYBR® Green) and qPCR samples which were either strongly positive for one assay or tested positive with more than one qPCR assay, a comparison was made as per Section 10.1.5. No significant relationship was seen, although there was a trend towards a positive relationship in the pellet samples (data not shown).

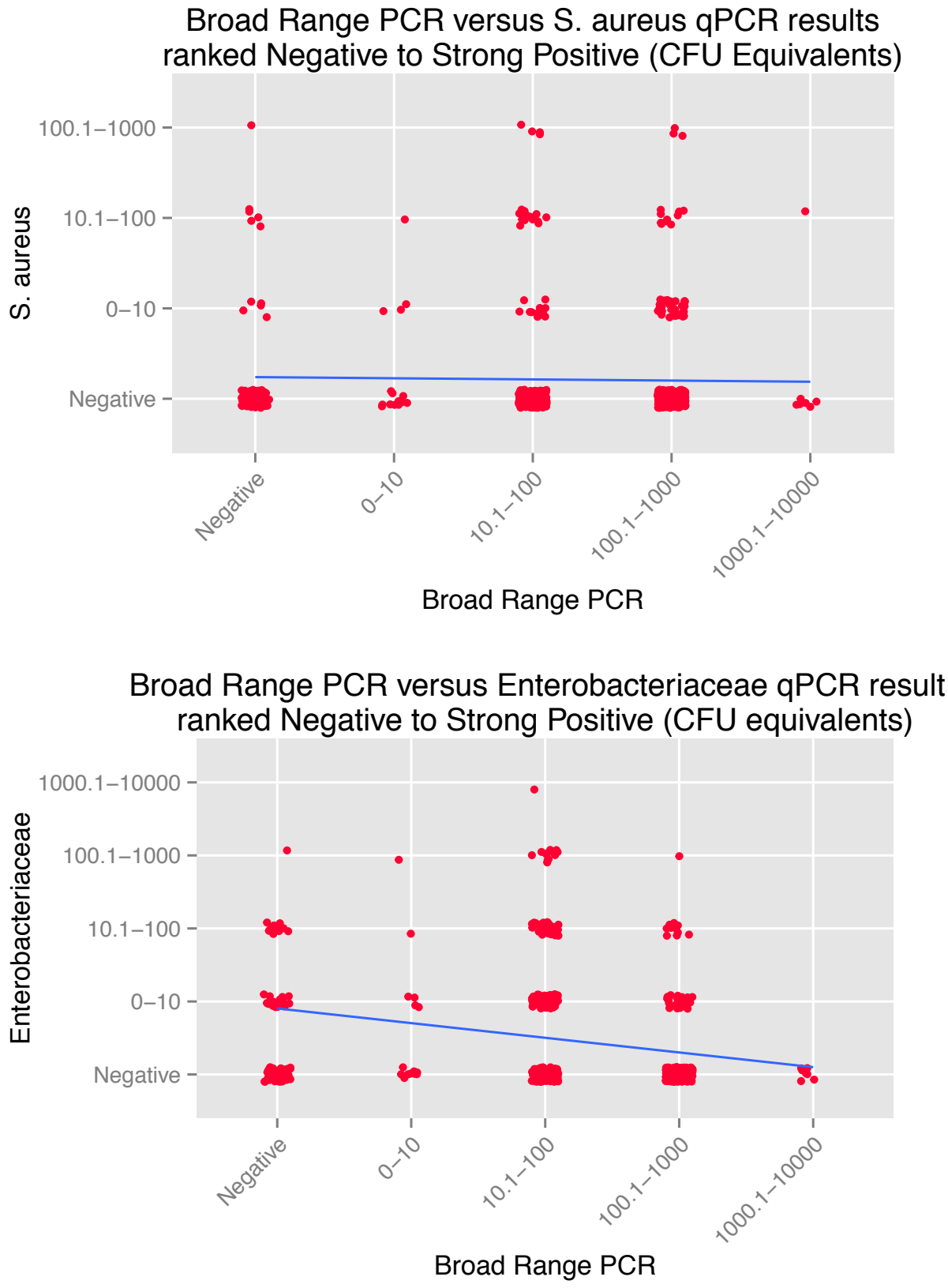


Figure 34 Scatterplots of qPCRs compared with broad range 16S rDNA PCR (high number of positive qPCRs), comparing ranking of negative to high positives

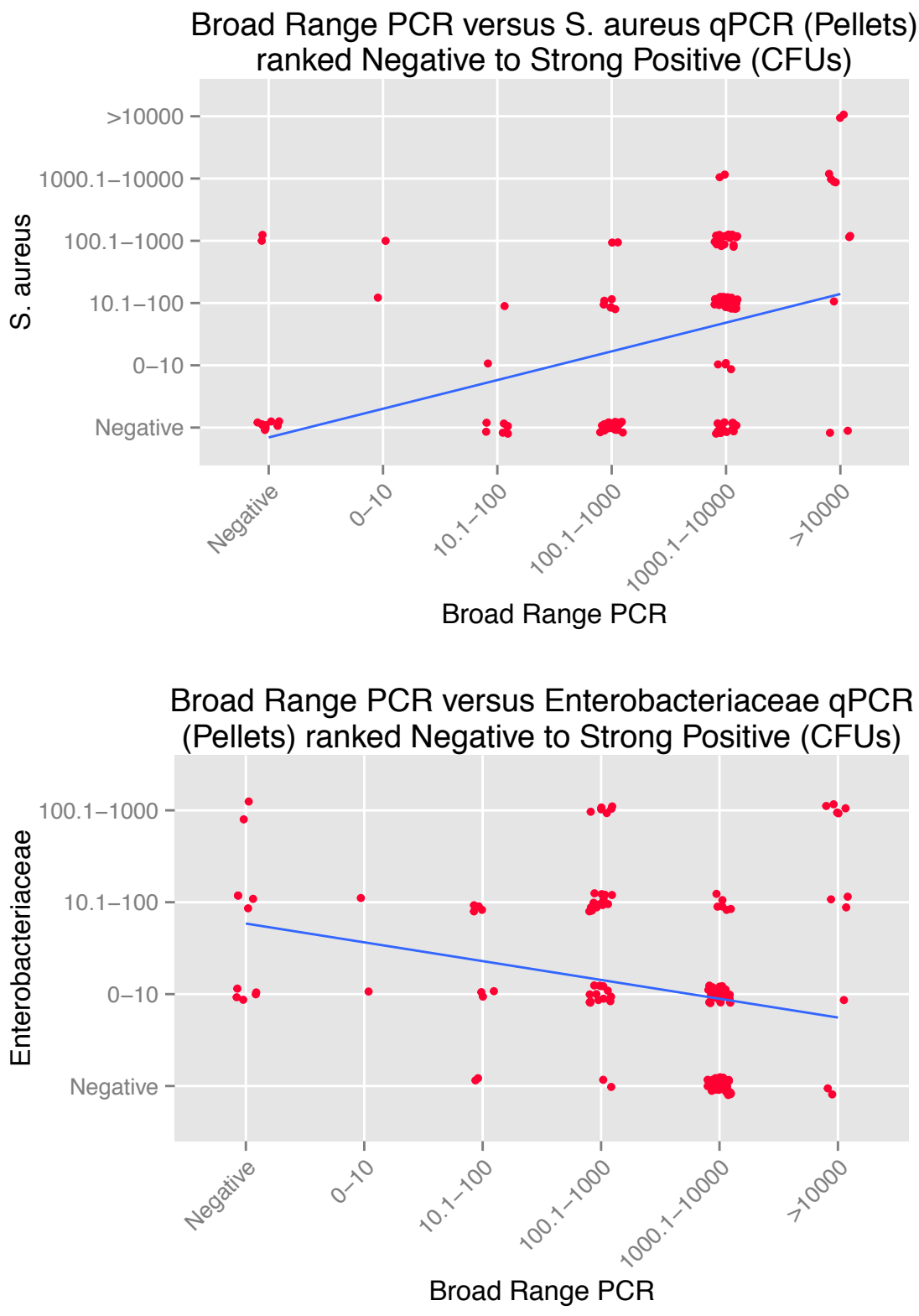


Figure 35 Comparison of qPCR versus broad range 16S rDNA PCR for pellet samples, comparing ranking of negative to high positives

13.5 Discussion

Interpreting the results from the broad range 16S rDNA PCR (SYBR® Green) assay in this study presents some challenges. Broad range 16S rDNA PCR has been used with and without subsequent sequencing as a marker of microbial translocation in HIV and SIV infection[8,54,59-66]. In most previous studies, the quantification has been recorded as 16S rDNA copies/µL as opposed to CFU equivalents/PCR reaction in comparison with standards of known CFUs, and the ranking scheme of negative to high positives used here. A disadvantage of the former method is that as the 16S rRNA gene has variable copy numbers between bacterial species, presenting the result as a single figure could lead to over-interpretation, especially in the absence of downstream sequencing to identify bacterial species present. This means that differences of less than one log in copies/µL might be artifactual and due to the species of bacteria present as opposed to the quantities. Owing to this assay's vulnerability to contamination, care must be taken to ensure that results are presented with concomitant provisos that even between runs within the same study comparisons may be inaccurate. Hence I opted for a cautious interpretation of quantification by comparison with CFUs of a known standard as an estimate of bacterial load, and performed NGS to confirm the presence of bacteria likely to be associated with microbial translocation as opposed to artifact and contamination.

As a broad summary, it appears that there was minimal difference between results obtained from the cases and controls, or the cases over time (Figure 31, Figure 32). Interquartile ranges were very wide and overlapping for all groups and over time as seen in Figure 32. This could indicate that there is genuinely no difference in the levels of bacterial DNA detected in these samples, or could indicate that the assay has reached the limits of detection when used without sequencing. Due to concern about the assay's vulnerability to contamination, results were interpreted conservatively: the limit of detection was determined by the lowest CT value of negative extraction and negative control samples even if this was lower than the threshold indicated by

the standards. This means the sensitivity (5-50 CFUs/PCR reaction) was lower than for specific qPCRs and so there may have been some false negative samples.

In terms of external comparison, the sensitivity of the assay equates to a sensitivity of 80-800 copies/µL[67]. This is much lower than the sensitivity of 5 copies/µL quoted in Jiang *et al.* which is the method used in several subsequent papers [59,61,65,66,68]. The median levels in HIV infected individuals detected in these papers vary from 5-400 copies/µL, and in one study, adults with chronic HIV had levels between 500-9000 copies/µL[63]. With the exception of Chevalier *et al.,* the levels generated by the assay of Jiang *et al.* tend to be lower than in this study (converting from CFU equivalents to median copies/µL of 1600-4180), but it is important to note that straightforward comparisons are unlikely to be useful owing to variability in the assay and the contrasting quantification methods used. Also, the method described by Jiang *et al.* does not include sequencing, and uses 45 thermal cycles, more than used in this study[59]. An alternative method described by Kramski *et al.* uses a shrimp nuclease to decrease contaminant DNA prior to amplification[64]. This modified assay has a quoted sensitivity of 10 copies/reaction, and reported 500-2500 16S rRNA copies/mL in HIV-infected adults from Australia[64]. A further method (again with no sequencing) used in a cohort of children in Spain found medians of 630 and 3162 copies/µL in HIV uninfected and infected children respectively[69,70]. It can be seen there is considerable variation in levels generated by different assays, and in copy numbers detected even in using similar assays in different studies. In summary, considering the difference (or lack of difference) between groups within this study are likely to be more useful than external comparators, owing to different methods used. In the absence of an external assay calibrator or standard, the usefulness of interstudy comparison of quantitative data is limited.

The apparent negative relationship seen between the *Enterobacteriaceae* assay and the broad range 16S rDNA PCR (SYBR® Green) could be due to the fact that the *Enterobacteriaceae* assay is also vulnerable to contamination and so is similarly limited in terms of its sensitivity and that the apparent association is spurious (Figure 34). It may be that there is some form of sample inhibition that impacts on one assay more than the other, although this pattern did not extend to the *S.aureus* assay

(Figure 34), and internal positive control results did not indicate inhibition. Another explanation is that all PCR primers have inherent bias, and so the broad range primers used in the SYBR® Green assay could be more biased towards Gram positive bacteria such as *S.aureus*, and away from *Enterobacteriaceae.* Additionally, *Enterobacteriaceae* have variable copy numbers of the 16S rRNA gene, from 1 to 11 copies depending on the species[71]. Even within species there can be variation[72]. By contrast, the *DnaK* gene targeted by the *Enterobacteriaceae* qPCR is a single copy gene[73]. The broad range 16S rDNA PCR may therefore give less accurate quantification of *Enterobacteriaceae*. Furthermore, the *Enterobacteriaceae* PCR was designed for use in a clinical setting, and so targeted towards clinically significant bacteria such as *Escherichia coli*[15]. The 16S rDNA PCR primers, while also aiming to target clinically significant bacteria, are subject to more constraints due to the need to cover as broad a range of bacteria as possible. Therefore compromises have to be made such that for a diverse family such as *Enterobacteriaceae,* contaminant rather than biologically significant species may be preferentially sequenced.

A broad comparison was made between the broad range 16S rDNA PCR (SYBR® Green) and samples which were either strongly positive in one qPCR assay or positive with several qPCR assays for the purposes of hypothesis generation. This comparison was to investigate whether there was any similarity between the numbers of CFU equivalents detected by the broad range assay and across the panel of specific assays. With the plasma samples, no similarity was seen but there was a trend towards a relationship in the pellet samples. This is likely to be driven by the positive association seen between broad range assay results and the *S. aureus* assay (Figure 35).

In summary, the results for the bacterial load as detected by a broad range 16S rDNA PCR were very similar between HIV-infected groups and controls, and over time within HIV infected groups, with the one exception of the rural group having a significantly lower bacterial load at baseline. This suggests that the low levels of bacterial DNA as detected by this assay are unlikely to play a significant biological role in this setting. Although results tended to be higher than those detected in other studies, there is considerable variability both in methods used elsewhere and in

bacterial levels detected which undermines the usefulness of external comparison.

These results should be interpreted together with those of Chapters 12 and 14.

14 Chapter 14 Next Generation Sequencing subsequent to Broad Range 16S rDNA PCR to Identify Bacterial DNA involved in Microbial Translocation in HIV-infected and Uninfected children.

14.1 Background

As discussed in sections 8.4.6.1 and 8.4.6.2, without sequencing, it is difficult to determine whether results from broad range 16S rDNA PCR are artefact as the assay is vulnerable to contamination. NGS offers the most comprehensive and effective method to sequence the full range of bacterial species which may be present in a sample.

14.2 Study Population

See section 9.1

14.3 Methods

NGS was carried out according to section 9.7. Barcoded primers (Table 39) were used at a concentration of 0.2pmol/µL. Post library amplification, the samples were initially cleaned individually using AMPure XP beads (Beckman Coulter) and a plate magnet at a ratio of 0.8:1 beads to sample. They were then normalised and pooled using SequalPrep normalization plate kits (Invitrogen). Several further steps of cleaning using AMPure XP beads (Beckman Coulter) at a ratio of 0.8:1 were carried out for each plate as guided by quantification with an Agilent high sensitivity DNA kit (hsDNA) using a Bioanalyser, comparing the ratio of primer dimer to library (see section 10.2.2). The libraries only successfully sequenced when the ratio of amplicon to primer dimer was at least 3:1. Final library quantification was carried out using a Qubit 2.0 (Life technologies) and 10pM library loaded onto a MiSeq (Illumina) as per manufacturer's protocol for 500 cycle V2 kits with the addition of custom sequencing primers and a 10% PhiX spike. Sequences were demultiplexed and filtered and OTUs assigned using QIIME (v1.8.0)(section 19.3, Appendix B) and results were analysed using R Studio (v.0.99.896) using Phyloseq (v.1.10.0) as per section 9.7.

14.4 Results

Samples available for analysis were the same as in section 12.4. In total, 176/808 (22%) samples produced sufficient quantities of amplified 16S library to be successfully sequenced. 105/140 (75%) pellet samples were sequenced along with 71/668 (11%) plasma samples (Table 28). None of the 113 control samples from HIV-uninfected children were sequenced successfully as insufficient 16S library was produced.

Table 28 Table of distribution of samples successfully sequenced using NGS by ART group and sample time points

	Baseline Pellet	Baseline Plasma	Week 12	Week 72
ART naïve	83/118 (70%)	53/112 (47%)	3/112 (3%)	0/112 (0%)
95% CI	61-78	38-57	6-8	0-3
ART experienced	22/22 (100%)	6/22 (27%)	1/22 (5%)	0/22 (0%)
95% CI	85-100	11-50	1-23	0-15
Gulu (ART naïve)	NA	0/52 (0%)	0/49 (0%)	7/49 (14%)
95% CI		0-7	0-7	6-27

Eleven controls were included: two mock communities (one was run in duplicate), two other positive controls, one negative extract on each run and two other negative controls on each run. As the number of OTUs generated tended to be low; both rarefied and unrarefied results are presented. The results were dominated by Staphylococcus spp. and Enterobacteriaceae with these OTUs also appearing in the negative experimental controls in both runs (Figure 36, Figure 37). Similar OTUs appeared in all three ART groups as can be seen from the phylogenetic tree node labels (Figure 36, Figure 37). Furthermore there was no clustering of the negative control groups within principle coordinates analysis: rather there was representation of all groups across the plot regardless of ART status(Figure 38).

Some species included in the mock communities were from families missing from the sequencing results: *Actinomyces odontolyticus, Deinococcus radiodurans,* and *Rhodobacter sphaeroides.* On checking the alignment of the primers with the 16S

rRNA gene of these species, there was a good match between the primers and the relevant regions of the genes. All other species in the mock communities were from families identified in the sequencing.

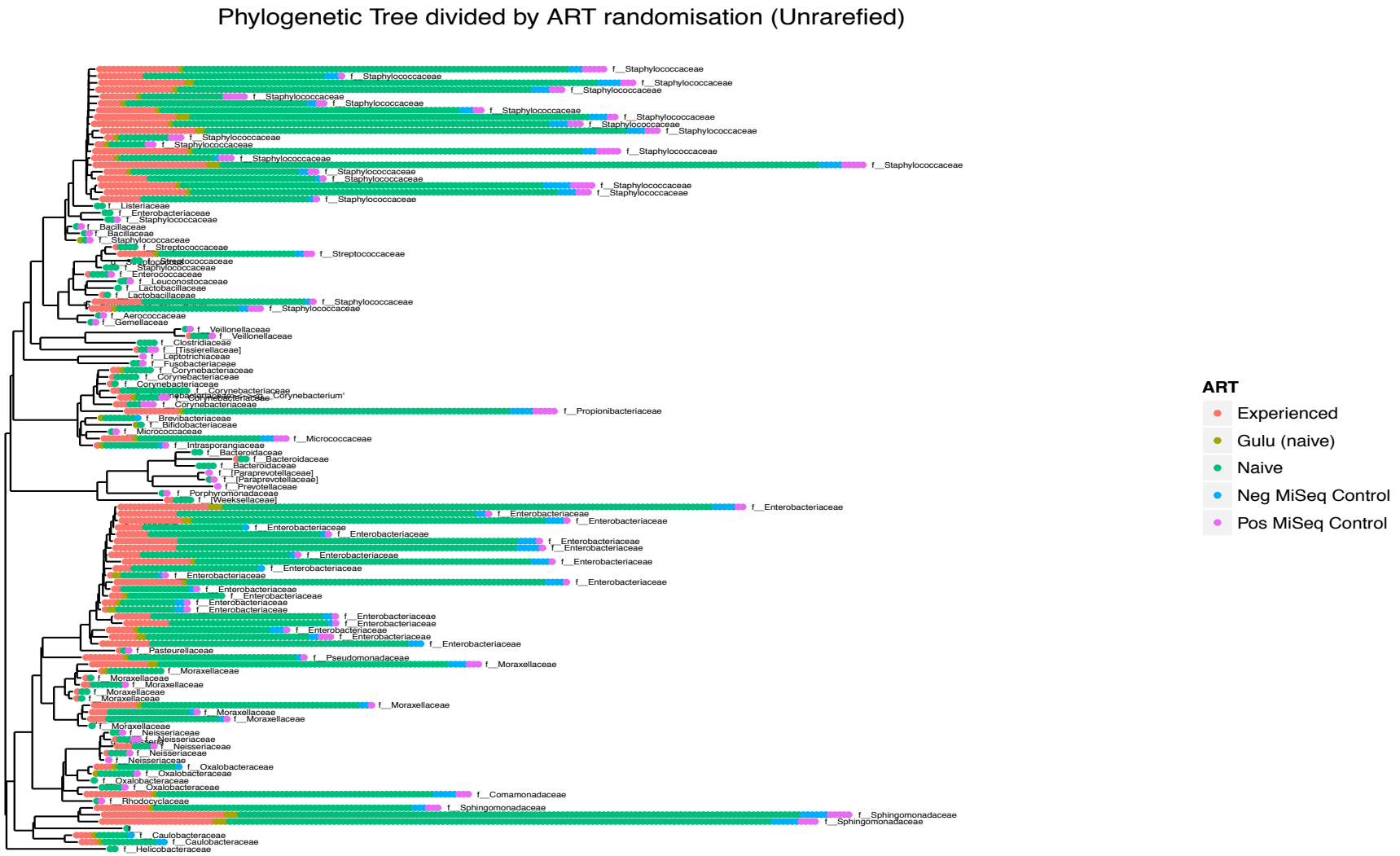


Figure 36 Phylogenetic Tree by ART randomisation (unrarefied): showing predominance of *Staphylococcacea, Enterobacteriaceae,* and *Sphingomonadaceae*. Negative experimental control samples (coloured blue) can be seen across most nodes

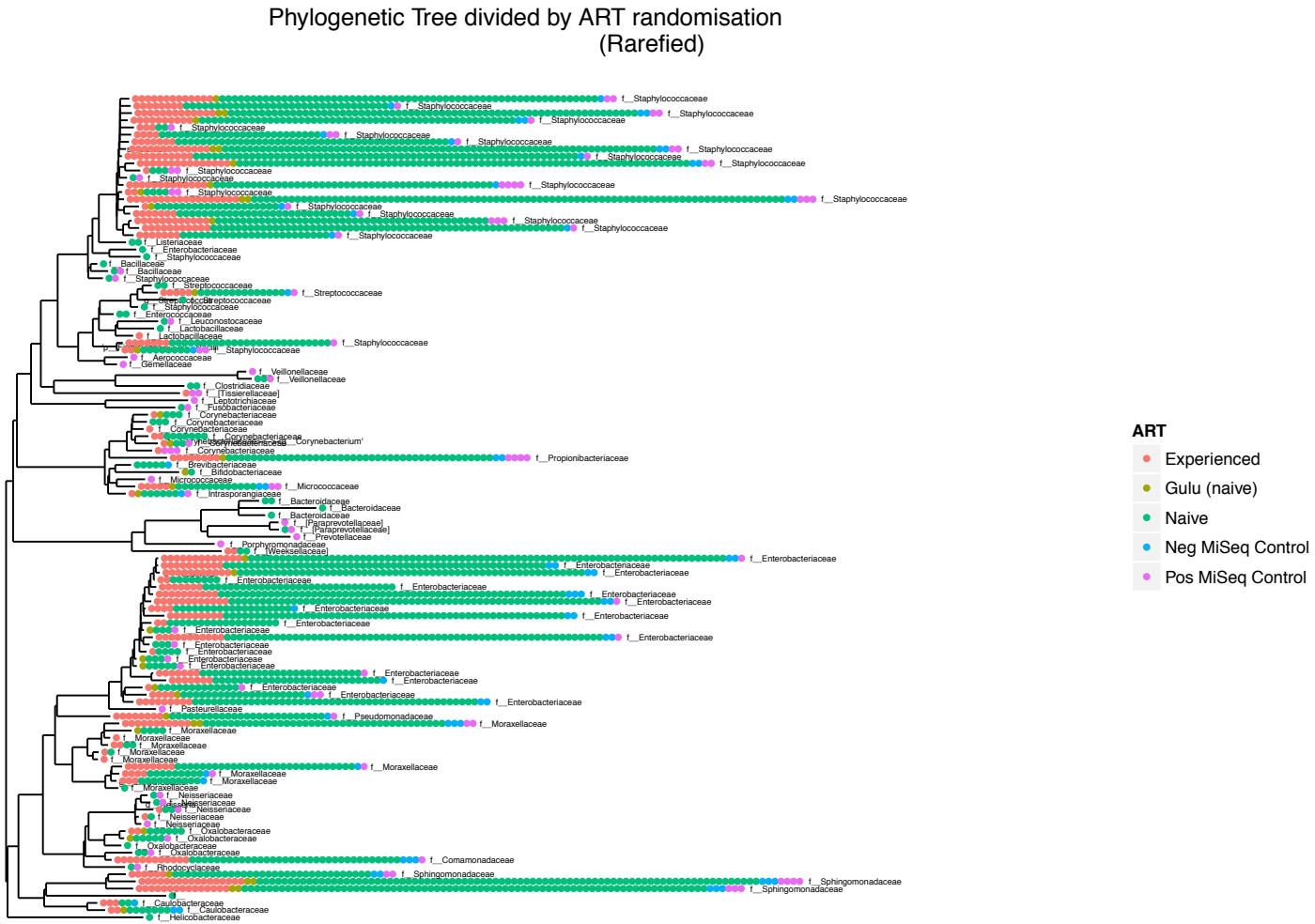


Figure 37 Phylogenetic Tree by ART Randomisation (rarefied) showing predominance of *Staphylococcacea, Enterobacteriaceae,* and *Sphingomonadaceae.* Negative experimental control samples (coloured blue) can be seen across most nodes

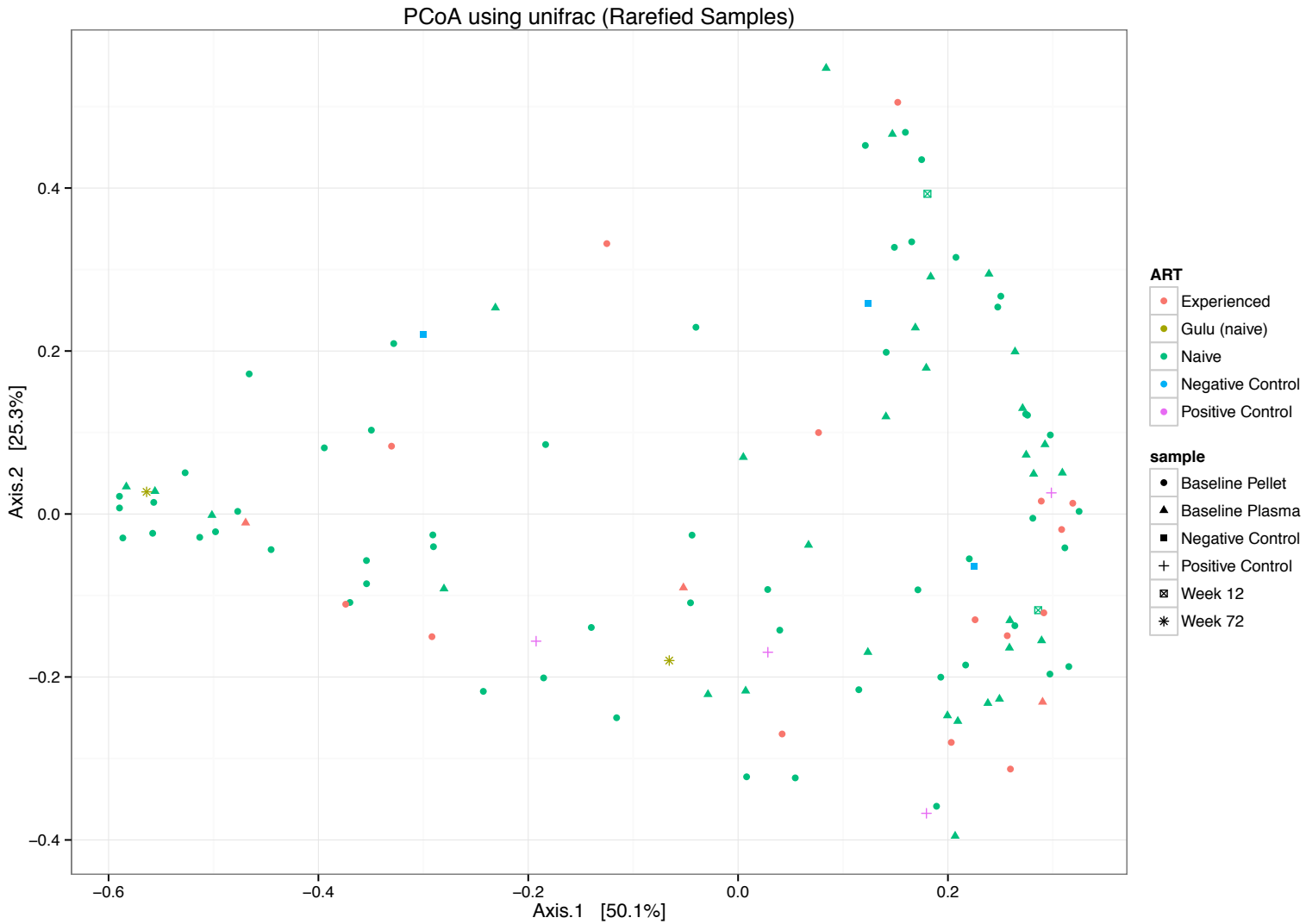


Figure 38 Principal Coordinates Analysis using Unifrac (rarefied) showing no specific clustering by sample type or ART randomisation. Negative experimental control samples (squares) can be seen dispersed across analysis with no specific clustering.

14.4.1 Removal of OTUs in Negative Control Samples

Examining the results using principal coordinates analysis (PCoA) (Section 9.9) demonstrated that the experimental controls were distributed across the results rather than clustered together (Figure 38). Therefore the OTUs contained within the controls were removed and the samples reanalysed both with and without rarefaction.

Families from which OTUs were removed were: Brevibacteriaceae, Caulobacteraceae, Comamonadaceace, Enterobacteriaceae, Intrasporangiaceae, Micrococcaceae, Moraxellaceae, Propionibacteriaceae, Pseudomonadaceae, Sphingomonadaceae, Staphylococcaceae and Streptococcaceae. After removal of experimental negative control OTUs, the resulting phylogenetic trees were scanty (Figure 39, Figure 40). Staphylococci and Enterobacteriaceae were still seen across both ART-naïve, ART-experienced and Gulu (ART-naïve) groups at low abundance (Figure 39, Figure 40). PCoA showed minimal clustering among ART randomisation groups or timings of sample (Figure 41).

OTUs present after removal of those seen in experimental negative control samples include: *Staphylococcaceae, Lactobacillaceae, Veillonellaceae, Clostridiaceae, Bacteroidaceae, Porphyromonadaceae, Enterobacteriaceae, Neisseriaceae, Moraxellaceae, Oxalobacteriaceae,* and *Helicobacteriaceae.*

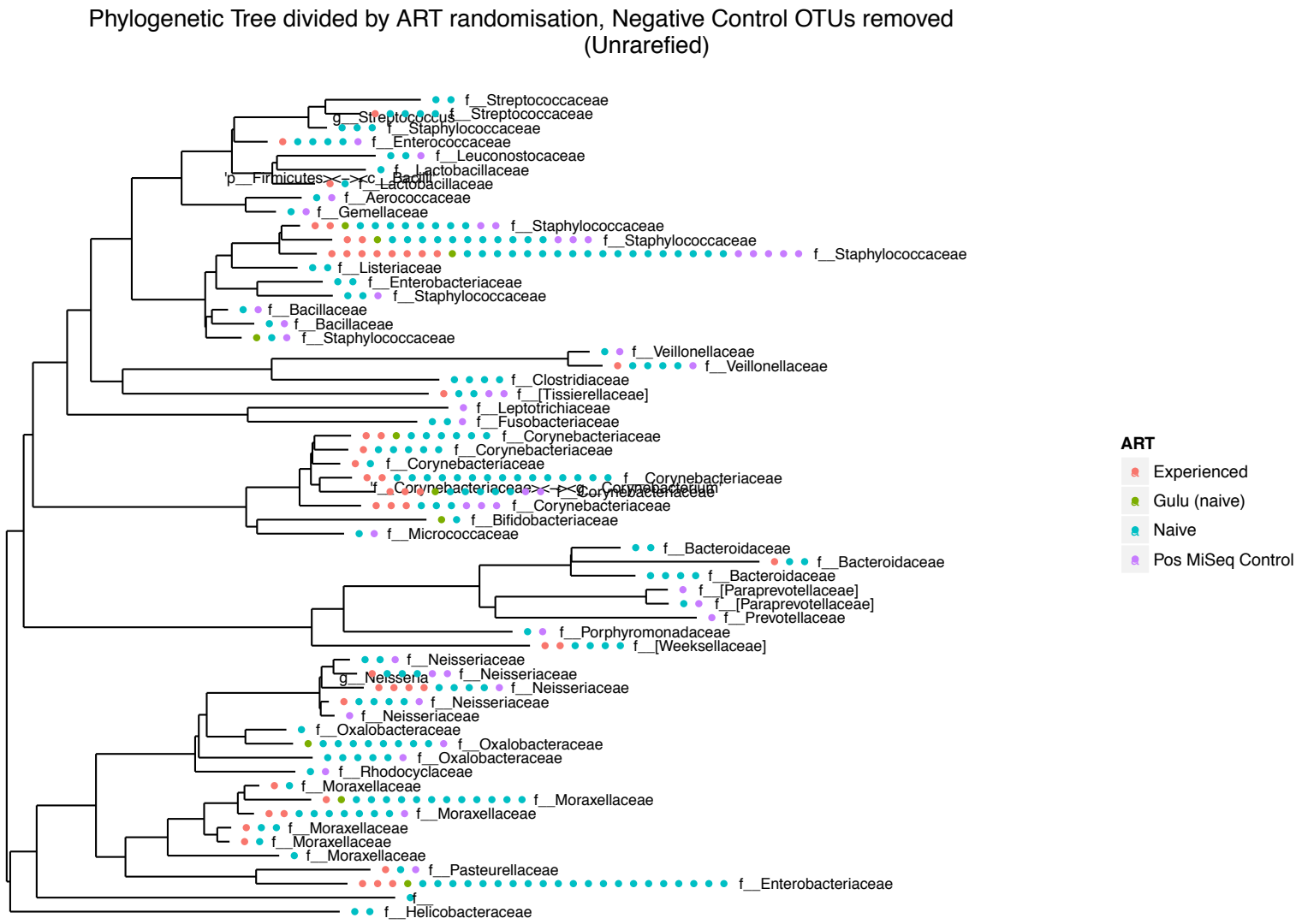


Figure 39 Phylogenetic tree by ART randomisation after removal of negative control OTUs (unrarefied), showing sparse OTUs including *Staphylococcaceae* amongst other families. OTUS derived from both ART-experienced and ART-naïve samples distributed across phylogenetic tree.

Figure 40 Phylogenetic Tree by ART randomisation, Negative Control OTUs removed (Rarefied) showing sparse OTUs including *Staphylococcaceae* amongst other families. OTUS derived from both ART-experienced and ART-naive samples distributed across phylogenetic tree.

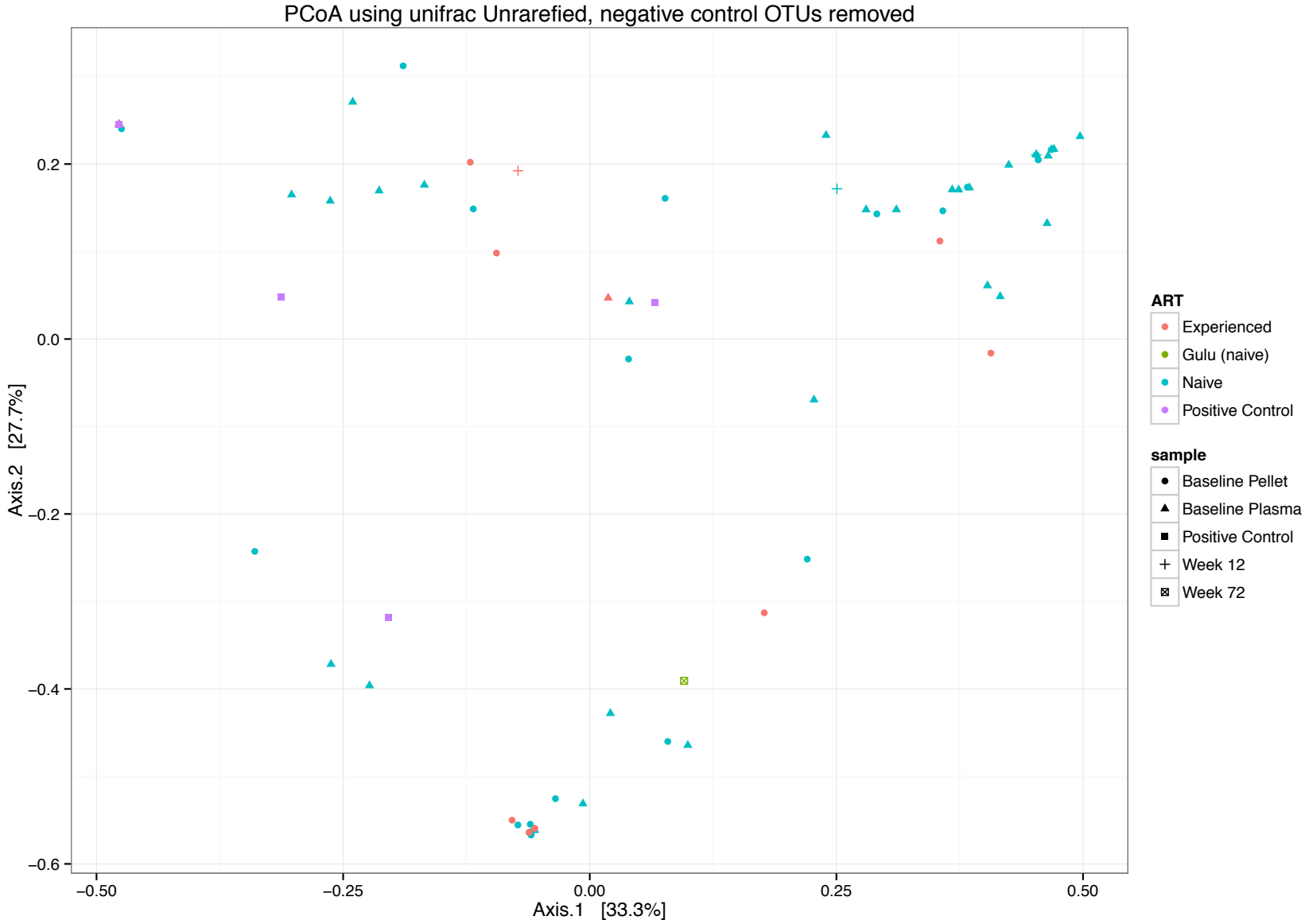


Figure 41 PCoA after removal of negative control OTUs (Rarefied): no clustering by sample type or timing. ART naïve and ART experienced samples distributed throughout.

14.4.2 Comparison of NGS Results with qPCR and Broad Range 16S rDNA PCR

The PCoA analysis using unifrac was repeated to look for associations with results from a) Broad range 16S rDNA PCR (SYBR® Green) and b) *S.aureus* and *Enterobacteriaceae* qPCR assays in both pellets and plasma samples. Minimal association was seen prior to removal of negative control OTUs (data not shown).

After removal of the negative control OTUs, there did appear to be some association in both the broad range 16S rDNA PCR (SYBR® Green) and the *S.aureus* PCRs in the pellet samples and clustering of the PCoA (Figure 42), in that more strongly positive samples clustered together, as did negative samples. However there was no distinct clustering seen between ART naïve and ART-experienced, and minimal relationship was seen in the plasma samples (Figure 43). There was no clustering seen between *Enterobacteriaceae* results and the PCoA from the next generation sequencing after removal of negative control OTUs (Figure 42, Figure 43).

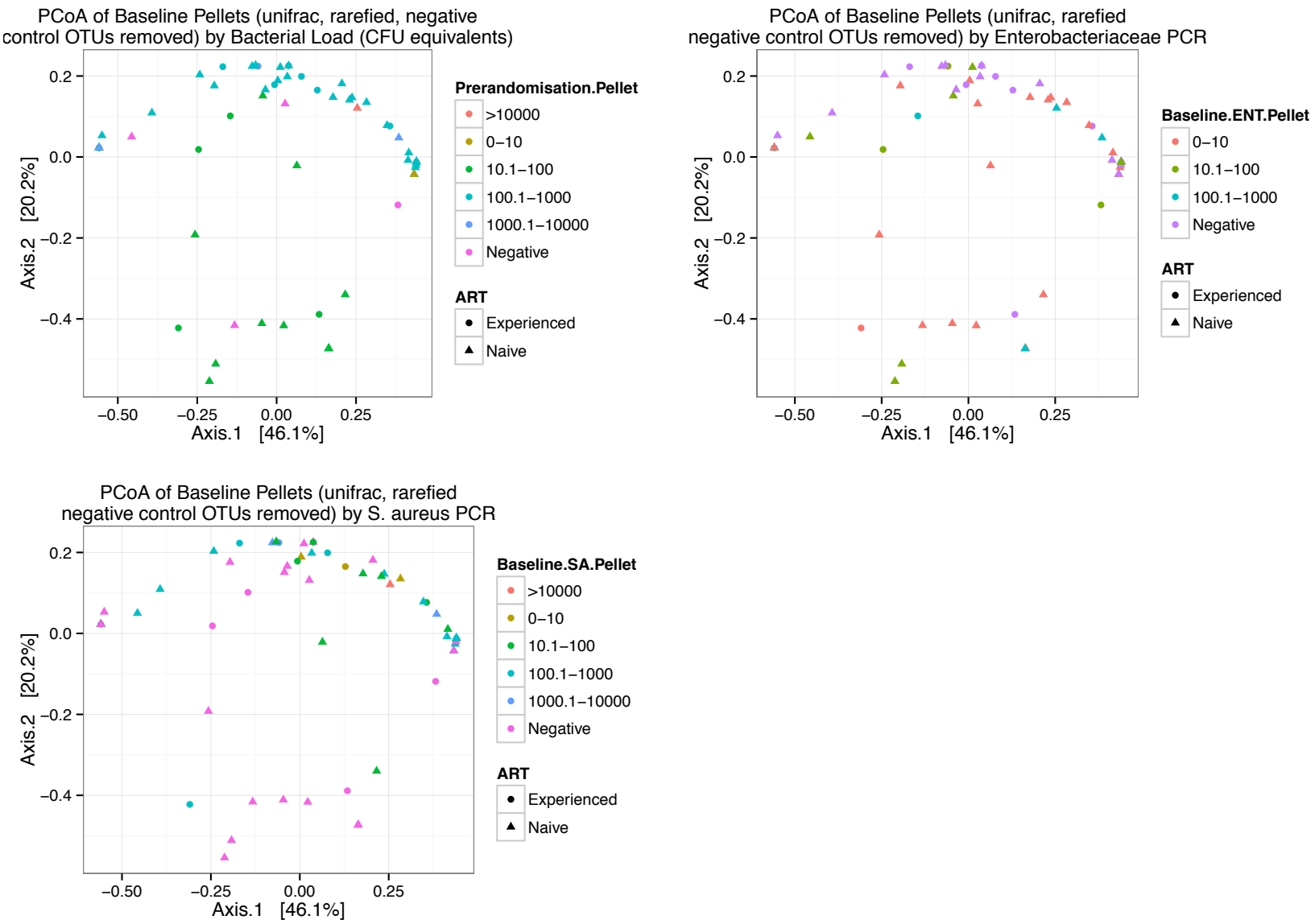


Figure 42 PCoA of Pellet Samples with Negative control OTUs removed compared with other PCRs: broad range 16S rDNA PCR, *Enterobacteriaceae* PCR and *S. aureus* PCR. Some clustering seen by negative and high results in broad range 16S rDNA PCR and *S.aureus* PCR, but not by negative or high results in *Enterobacteriaceae* PCR.

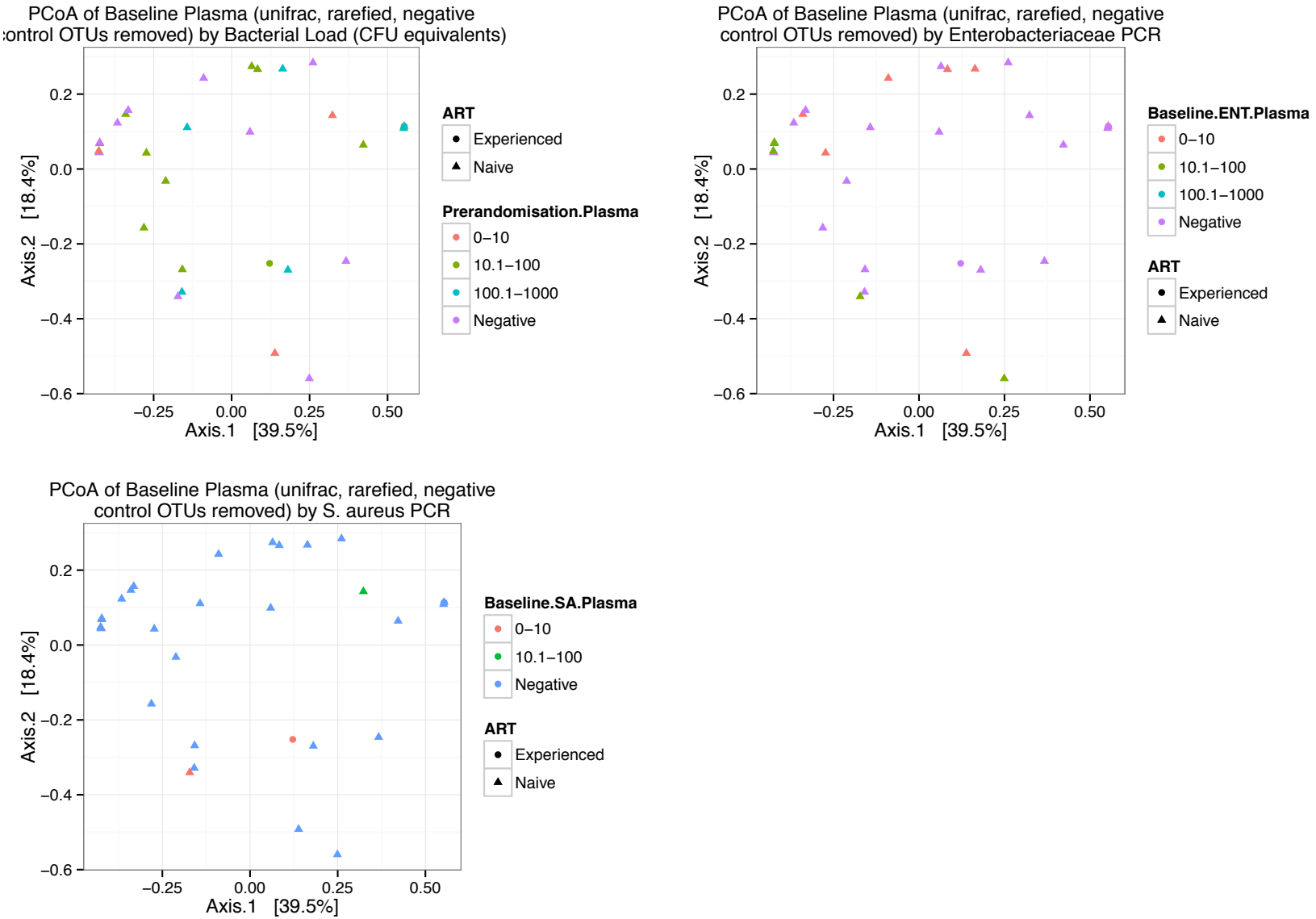


Figure 43 PCoA of Plasma samples after negative control OTUs removed compared by other PCR assays: broad range 16S rDNA PCR, *Enterobacteriaceae* PCR and *S. aureus* PCR. Minimal clustering seen by negative or high results in any of the three assays.

14.5 Discussion

This study has demonstrated the challenges of NGS when analysing samples with low bacterial load. The assay was working at the limit of detection, as seen by the fact that insufficient library was generated in many samples to sequence successfully, and even in the successful runs, many OTUs generated were also seen in negative control samples (Figure 36, Figure 37). On removing the OTUs seen in negative controls, the resulting OTUs generated a scanty phylogenetic tree both with and without rarefaction (Figure 39, Figure 40).

That many OTUs were seen in negative control samples highlights that 16S rDNA PCR is vulnerable to contamination and the crucial importance of sequencing to indicate whether results generated are artefact. Of previous studies using 16S rDNA PCR to investigate microbial translocation in the setting of HIV/SIV infection, only two have used any form of sequencing on plasma samples[8,54]. The Sanger sequencing method used in these two studies revealed *Enterobacteriaceae* and *Lactobacillaceae* which could be implicated in microbial translocation as well as families such as *Comamonadaceae,* and members of the *Enterobacteriaceae* family such as *Rahnella sp.* which could be contaminants.

The pilot study for this thesis used NGS (Roche 454 platform®) and identified families in the blood of HIV-infected children in London that could be consistent with microbial translocation[9]. These included *Veillonellaceae* and *Lactobacillaceae.* However, the numbers of samples was low and no control groups were available. It is possible that the results of the pilot study also could be consistent with contamination. The only other study to have used next generation sequencing (also using the Roche 454 platform®) involved investigating bacterial DNA found in the colon, liver and mesenteric lymph nodes of SIV-infected RMs which identified certain families within the phylum of Proteobacteria to translocate preferentially compared with others, and members of the Bacteroidetes phylum[62]. Interestingly though, the families identified to have preferentially translocated to lymph nodes included

Sphingomonadaceae, Caulobacteraceae, and *Comamonadaceace,* all of which were identified in negative experimental controls in this study and removed. It is challenging to identify which species might be genuine indicators of microbial translocation and which represent artefact.

Within this study, of the families that were identified, some might be implicated in gut translocation, namely *Enterobacteriaceae* (as seen in previous studies[8,59]), *Clostridiales, Veilonella, Lactobacillales* and *Bacteroides,* but these were seen at low levels and represented both in ART-naïve and ART-experienced groups. Other families could be implicated in translocation from sites other than the intestine, such as the nasopharynx (*Neisseriaceae, Moraxallaceae, Corynebacteriaceae* and *Porphyromnodaceae)*, although some species are also represented in mouth flora (*Porphyromnodaceae* and *Fusobacterium).* The *Staphylococcaceae* and *Streptococcaceae* are difficult to interpret as they could have been derived from the skin, therefore representative of contamination, but are also seen in gut flora. Taken altogether, as these families are seen in low levels in both ART-experienced and ART-naïve children, this could be representative of baseline low levels of ubiquitous translocation that may be seen in healthy individuals, rather than a phenomenon that is unique to, or at least more severe in those with uncontrolled HIV[62,74-77].

An important limitation of this study is that some species from the mock communities were not identified by sequencing. This may have been due to issues with the mock communities themselves as the same species were not identified in the mock communities in a parallel project with different samples[78]. It should be noted that in any case, two of the three families from which species were omitted, namely *Rhodobacter* and *Deinococcaceae* are more usually associated with soil and are not well known gut commensals or pathogens[78-80]. The omission of *Actinomyces odontolyticus* is more concerning as this is a recognised pathogen and can be found in the human gut[81]. However, otherwise coverage of the mock community species was reasonable.

Primer bias is unlikely to be the cause of the omission of the three mock community species as the primer sequences were checked relative to the 16S rRNA genes of the species in question. However, primer bias could apply to species present in experimental samples that have also not sequenced successfully. This is an issue with any PCR-based reaction which is that all primer sets can be biased either positively or negatively against particular bacterial strains. Use of alternative primer sets such as the qPCR assays in this study should go some way to addressing this, but these assays were also working at the limit of detection and are narrow in their targets.

It is also important to note how pellet samples were more likely to sequence successfully than plasma samples, a pattern also seen in the PCR results in sections 12 and 13. This meant that both the Gulu group, the HIV-uninfected control group and the samples at different time points when no pellet samples were available were less likely to have contributed data to the sequencing. However, 47% (95% CI 38-57) of the ART-naïve group and 27% (95% CI 11-50) of the ART experienced group had plasma samples that successfully sequenced at baseline, versus 3% (95% CI 6-8) and 5% (95% CI 1-23) respectively at week 12, and none at week 72. Also, 14% (95% CI 6-27) of the Gulu samples at week 72 sequenced successfully but none at any other time point.

This could be consistent with genuine microbial translocation rather than contamination. However if this were so, a difference between the ART groups would be predicted. It is more likely that there may have been some difference in the method of sample collection and storage at these time points, particularly with regards to the Gulu samples. The Gulu samples had to undergo two stages of transport, firstly from Gulu to Kampala and then on to London, so increasing the chances of problems in the cold chain. It could also be the result of a batch effect. In order to choose samples most likely to successfully sequence, they were batched according to bacterial load by 16S rDNA PCR (SYBR® Green) result as per section 10.2.2. This meant that for the most part, samples were not batched together by site or by time point, but some underlying effect cannot be ruled out.

Considering why pellet samples might be more likely to yield higher quantities of DNA across all three assays, this might be a true representation of larger quantities of bacterial DNA present in the sample, potentially as a result of phagocytosed intracellular bacterial DNA in host cells being spun into the pellet. As the bacterial families identified in pellets and plasma were similar by qPCR and NGS (although the levels differed between sample types), it appears unlikely that different species were preferentially phagocytosed as hypothesised *a priori.* Furthermore, although the higher levels in pellet samples might be due to translocating bacteria within phagocytes, the results might also be due to host DNA acting as carrier DNA, increasing the proportion of recovered bacterial DNA from the sample. The quantity of host DNA has been demonstrated in previous studies to be important in successful recovery of microbial DNA in PCR assays[82]. It may also be consistent with contamination at collection, as any bacteria introduced may be concentrated within phagocytes or increased in pellets due to centrifugation

Although the OTUs found in the NGS results (largely *Staphylococci* and *Enterobacteriaceae*) support the results of the qPCR assays, there was minimal clustering on PCoA analysis by high or low results according to these assays prior to removal of the OTUs from negative controls (data not shown). After removal of the negative control OTUs, this pattern was still seen both with the plasma samples versus all other assays, and with the pellet samples for the *Enterobacteriaceae* PCR (

Figure 42, Figure 43). This indicates that there was not a consistent association between samples with high numbers of *Enterobacteriaceae* OTUs and those with a high result for *Enterobacteriaceae* on qPCR. This again indicates that these assays were at the limits of detection, and results were susceptible to contamination and variability.

There was some clustering of high *S.aureus* assay results and also with the broad range 16S rDNA PCR (SYBR® Green) results for the pellet samples (

Figure 42) after removal of the negative control OTUs. However, there was no clustering around ART groups. Therefore, it is difficult to extrapolate biological

significance from this result, as these OTUs were identified in both ART-naïve and ART-experienced children. Also, even in the pellet samples, these OTUs were seen at very low levels. It is possible therefore that they are the result of artefact, whether during sample collection (e.g. from the skin of children) or from the reagents (though this is less likely given that OTUs seen from experimental negative controls have been removed).

In summary, there was no consistent difference seen between OTUs generated from samples from ART naïve or ART experienced children, and overall, the number of OTUs generated were extremely low. Although some species detected by NGS might have been gut derived, the levels identified, particularly after removal of negative control OTUs, were very low. It is difficult conclude with confidence that these results are not artefact as the assay was working at the limit of detection.

15 Chapter 15 Assessment of Gastrointestinal Tract Injury and Anthropometry in relation to Molecular Markers of Microbial Translocation in HIV infected and Uninfected children.

15.1 Introduction

As per section 8.4.7 I-FABP was tested in all plasma samples available in this study in order to investigate differences in HIV-negative versus HIV positive children, changes over time on therapy in those with HIV and between HIV positive children living in urban (JCRC Kampala) versus rural (Gulu) environments. Anthropometric measures were compared with molecular assay results in order to investigate differences in between groups and changes over time on therapy in those with HIV.

15.2 Study Population

Please see section 9.1

15.3 Methods

Assays were carried out on all available plasma samples as per section 9.8. Calculations were carried out to compare anthropometric measures with assay results including I-FABP, broad range 16S rDNA PCR results and qPCR results (as per Table 10) at baseline and week 96 for HIV-infected groups. Analysis was as per section 9.10.

15.4 Results

For the I-FABP assay, at baseline, 109 results were available for the ART-naïve group, 20 for the ART experienced and 46 from Gulu. For the naïve control group, 89 results were available and 20 from the experienced control group. At week 12, 107 were available for ART naïve, 19 for ART experienced and 39 from Gulu. Eleven samples were available from week 24, 8 from ART naïve and 3 from Gulu. At week 72, 111 ART-naïve results, 21 ART-experienced and 39 Gulu results were available.

Missing results at baseline were as follows: ART-naïve 8%, 9 samples missing and 1 assay failure; ART-experienced 9%, 2 samples missing; Gulu 16%, 8 samples missing, 1

assay failure; control groups- none missing and no assay failures (Appendix D Missing samples across groups and time points). At week 12, 10% ART-naïve were missing results (11 missing samples, 1 assay failure), 14% ART experienced (all missing samples) and 19% from Gulu, (15 missing samples, 1 assay failure). At week 72, 3% ART naïve (all missing samples), 5% ART-experienced (missing samples) and 22% Gulu samples were missing (10 missing samples, 2 assay failures).

Gulu samples were significantly more likely to be missing than ART-naïve or ART-experienced groups ($p<0.001$ for both). Age and baseline CD4 T cell count were similar in the groups for whom data were missing and those who had complete data.

Baseline anthropometric measurements were available for all children except one from the naïve control group. In the HIV infected groups, anthropometric measurements at week 96 were available for 108/120 (90%) ART-naïve children (urban), 21/22 (95%) ART-experienced children and 40/55 (73%) from the Gulu group (rural, ART-naïve) (Figure 22). For the comparisons with molecular test results, I compared three anthropometric measures (weight-for-age, height-for-age and BMI-for-age z-scores) with three molecular tests (broad range PCR, *S.aureus* and *Enterobacteriaceae* qPCR) at three time points in three groups, at baseline between five groups and between pellets and plasma between two groups, meaning 3x3x3x3+5x3x3+2x3x3= 153 tests. I would therefore expect there to be 7-8 tests to be positive at p=0.05 by chance, 1-2 tests at p=0.01 and <1 at p=0.001.

15.4.1 I-FABP over time by ART Group

At baseline, I-FABP was lower in the ART-naïve and ART-experienced groups than both control groups and the Gulu (rural) groups ($p<0.001$ for ART-naïve compared with controls and Gulu, $p=0.041$ for ART-experienced versus controls) (Table 29, Figure 44). By week 12, the levels of I-FABP in the ART naïve group had increased to the point that they were no longer different from controls ($p=0.45$) (Spearman correlation coefficient of increase in ART naïve: 0.19, $p=0.05$) but were still significantly lower than the Gulu group ($p=0.009$)(Table 29, Figure 45). The I-FABP levels in the ART-experienced group had also increased to be no longer significantly

different from controls (p=0.97, Spearman correlation coefficient versus baseline 0.46, p=0.05)(Table 30). By Week 72, the levels in the ART experienced group had continued to increase so as to be significantly higher than in the ART-naïve group (p=0.0003, Spearman correlation coefficient 0.55 compared with baseline, p=0.01). The levels in the Gulu group were also significantly higher than in the JCRC (urban) ART-naïve group (p=0.0001), but the levels in the Gulu (rural) group did not change significantly over time (Figure 45, Table 30). The rate of increase in I-FABP in the ART naïve group slowed by week 72 (median at week 12 140 pg/mL, IQR 82-249; median week 72 133 pg/mL, IQR 83-193). Subgroup analysis of those in the 1st quartile (I-FABP <83 pg/mL at week 72) revealed no significant relationship between CD4 percentage at baseline or week 96, age or HIV viral suppression at week 96.

Table 29 I-FABP by ART group over time (comparisons using unpaired Wilcoxon rank sum test)

	Naïve	Experienced	P value naïve versus experienced	Naïve Controls	Experienced Controls	P value cases versus controls	Gulu	P value (Gulu versus naïve cases)
Baseline I-FABP (pg/ml) (Median, IQR)	74 40-125 n=109	72 53-111 n=20	0.79	140 75-222 n=89	160 65-305 n=20	**<0.0001 (naïve)** **0.041 (exp)**	249 189-296 n=46	**<0.001**
W12 I-FABP (pg/ml) (Median, IQR)	140 82-249 n=107	111 91-299 n=19	0.94	NA	NA	0.45 (naïve) 0.97 (exp)	222 137-279 n=39	0.009
W24 I-FABP (pg/ml) (Median, IQR)	169 75-263 n=8	n=0	NA	NA	NA	0.62 (naïve) 0.19 (exp)	55 55-72 n=3	0.19
W72 I-FABP (pg/ml) (Median, IQR)	133 83-193 n=111	207 153-323 n=21	**0.0003**	NA	NA	0.87 (naïve) 0.10 (exp)	199 154-296 n=39	**0.0001**

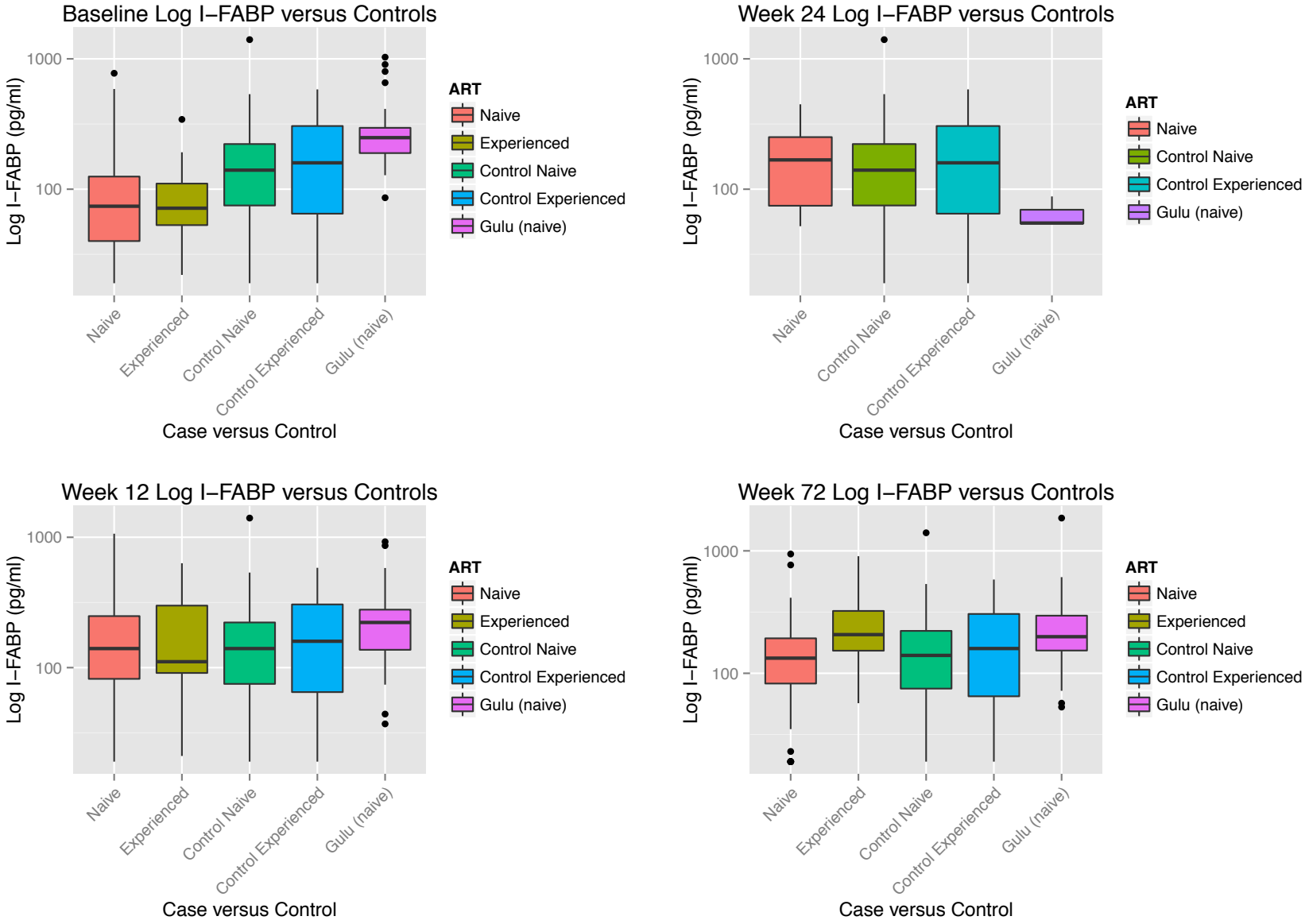


Figure 44 Boxplots of I-FABP over time by ART group.

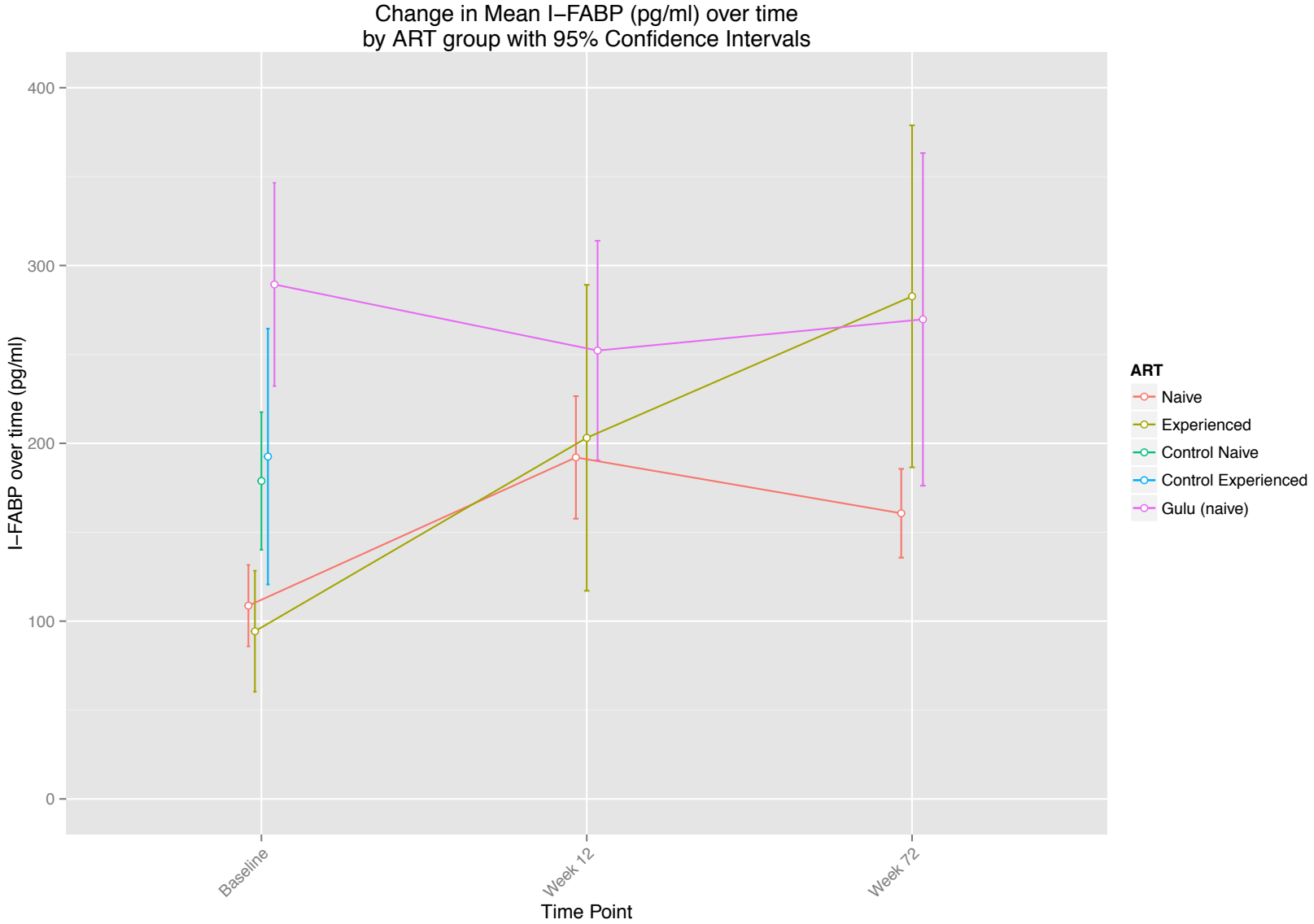


Figure 45 Change in Mean I-FABP over time by ART group with 95% Confidence Intervals

Table 30 Spearman Correlation coefficients comparing I-FABP at different time points by ART group

Spearman Correlation using complete observations		ART Naïve		ART Experienced		Gulu (naïve)	
		Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
PreRandomisation	Week 12	0.19	0.05	**0.46**	**0.05**	0.20	0.26
PreRandomisation	Week 24	-0.9	0.85	**NA**	**NA**	1	1
PreRandomisation	Week 72	0.02	0.87	**0.55**	**0.01**	0.01	0.97
Week 12	Week 72	-0.01	0.95	**0.51**	**0.02**	-0.13	0.49

15.4.2 I-FABP compared with Broad Range 16S rDNA PCR (SYBR® Green)

Results obtained with I-FABP assay were compared with the bacterial load (CFU equivalents) generated by the broad range 16S rDNA PCR using SYBR® Green. There did appear to be a borderline significant negative relationship between I-FABP and the PCR result in the ART-experienced group at week 12 (Spearman correlation coefficient -0.47, p=0.04)(

Figure 46, Table 31). No other significant relationships were seen in any ART group at other time points.

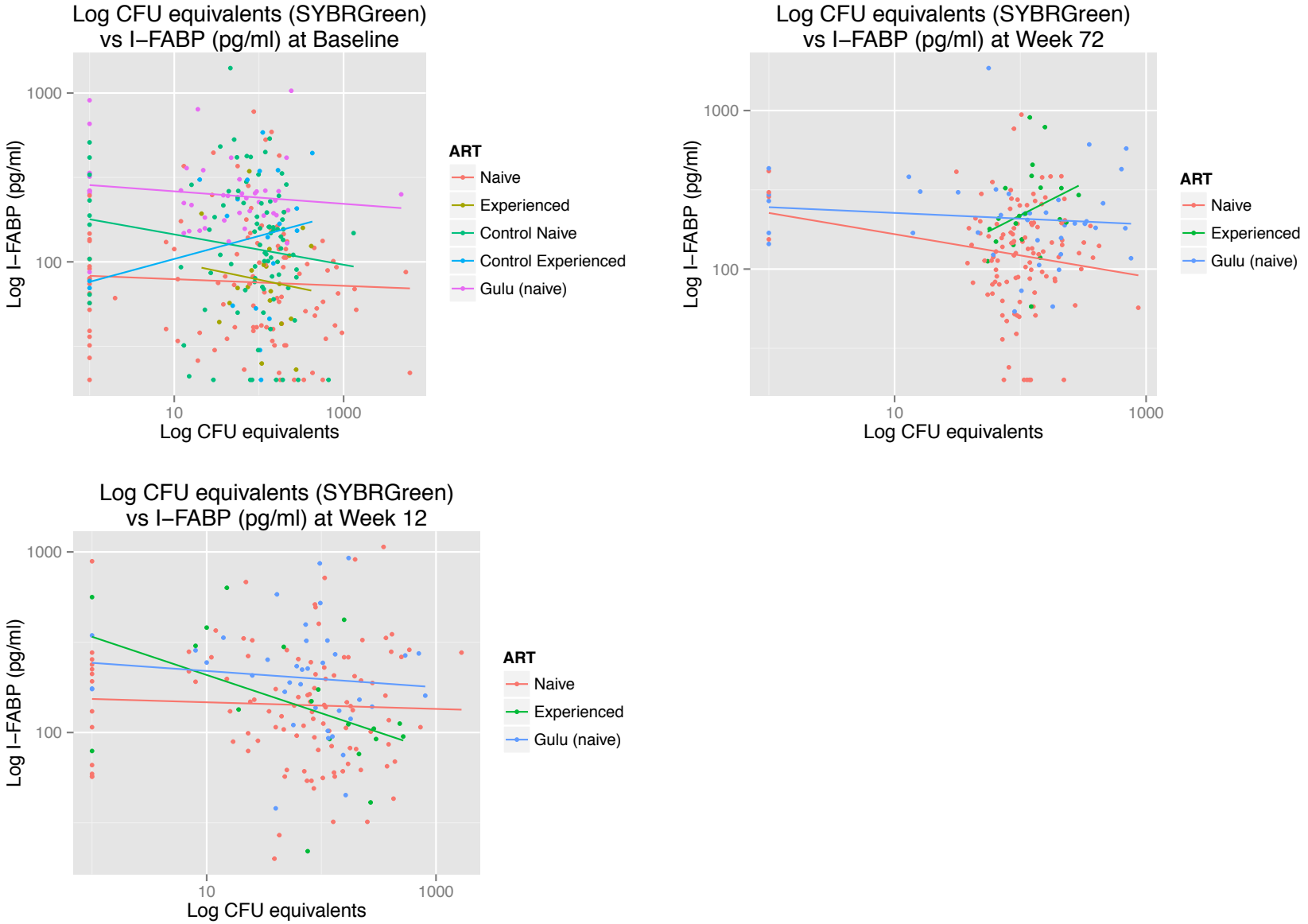


Figure 46 Scatterplot comparisons of I-FABP and bacterial load (broad range 16S rDNA PCR using SYBR® Green) by ART group at different time points

Table 31 Spearman correlation coefficients between I-FABP and Bacterial Load (CFU equivalents by broad range 16S rDNA PCR) by ART group over time.

Spearman Correlation using complete observations		ART Naïve		ART Experienced		Naïve Controls		Experienced Controls		Gulu (naïve)	
		Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
Baseline I-FABP	Baseline CFU	-1.1	0.26	-0.02	0.95	-0.18	0.09	0.21	0.38	-0.19	0.22
Week 12 I-FABP	Week 12 CFU	-0.02	0.81	**-0.47**	**0.04**	NA	NA	NA	NA	-0.22	0.18
Week 24 I-FABP	Week 24 CFU	0.24	0.57	NA	NA	NA	NA	NA	NA	0.5	1
Week 72 I-FABP	Week 72 CFU	-0.04	0.64	0.42	0.06	NA	NA	NA	NA	-0.11	0.51

15.4.3 Broad Range 16S rDNA PCR Results versus Anthropometric Measures

Baseline and week 72 bacterial load (CFU equivalents generated by broad range 16S rDNA PCR using SYBR® Green) results were compared with baseline and week 96 anthropometry measurements by ART group (

Figure 47,

Figure 48, Table 32). At baseline, there was a significant positive correlation between the BMI-for-age-z score and bacterial load (CFU equivalents) in the Gulu group (rho=0.28, p=0.05), and no other significant correlation seen in any other group. At week 72, the positive relationship between BMI-for-age-z score and bacterial load (CFU equivalents) in the Gulu group persisted (rho=0.36, p=0.03). There was also a significant negative relationship between the height-for-age z-score in the ART-naïve group and bacterial load (CFU equivalents) at week 72 (rho=-0.21, p=0.03).

Considering the pellet samples, there was no significant relationship between the bacterial load (CFU equivalents) and weight-for-age z-scores in either the ART-naïve group at baseline (rho=0.05, p=0.63), or the ART-experienced group (rho=-0.24, p=0.29).

Similarly there was no relationship between bacterial load (CFU equivalents) and BMI-for-age z-scores (ART-naïve rho=0.01, p=0.89; ART-experienced rho=0.21, p=0.34) or height-for-age z-scores (ART-naïve rho=0.02, p=0.82; ART-experienced rho=0.36, p=0.1).

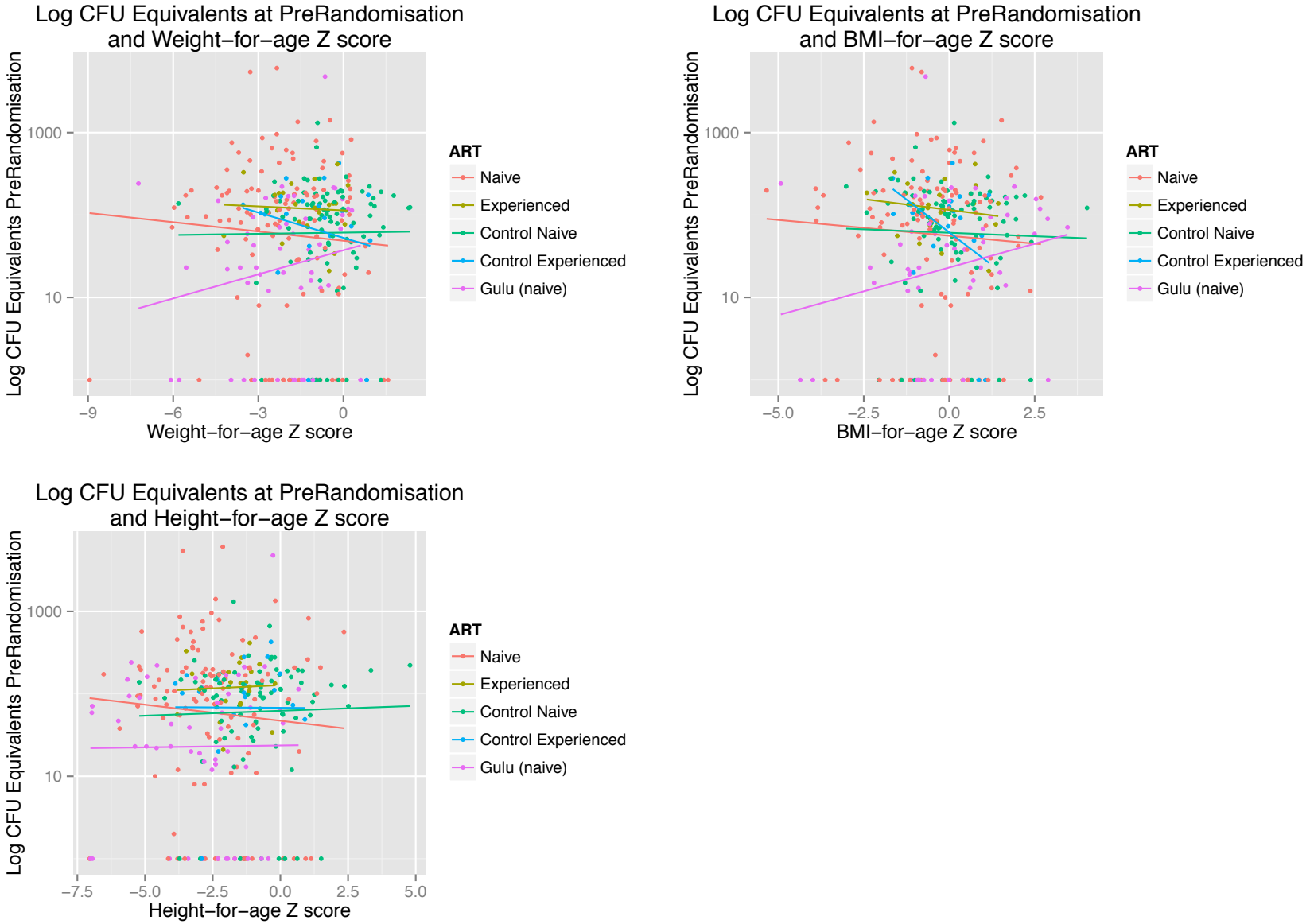


Figure 47 Scatterplot of bacterial load (CFU Equivalents generated by broad range 16S rDNA PCR using SYBR® Green) compared to anthropometric measurements at Baseline

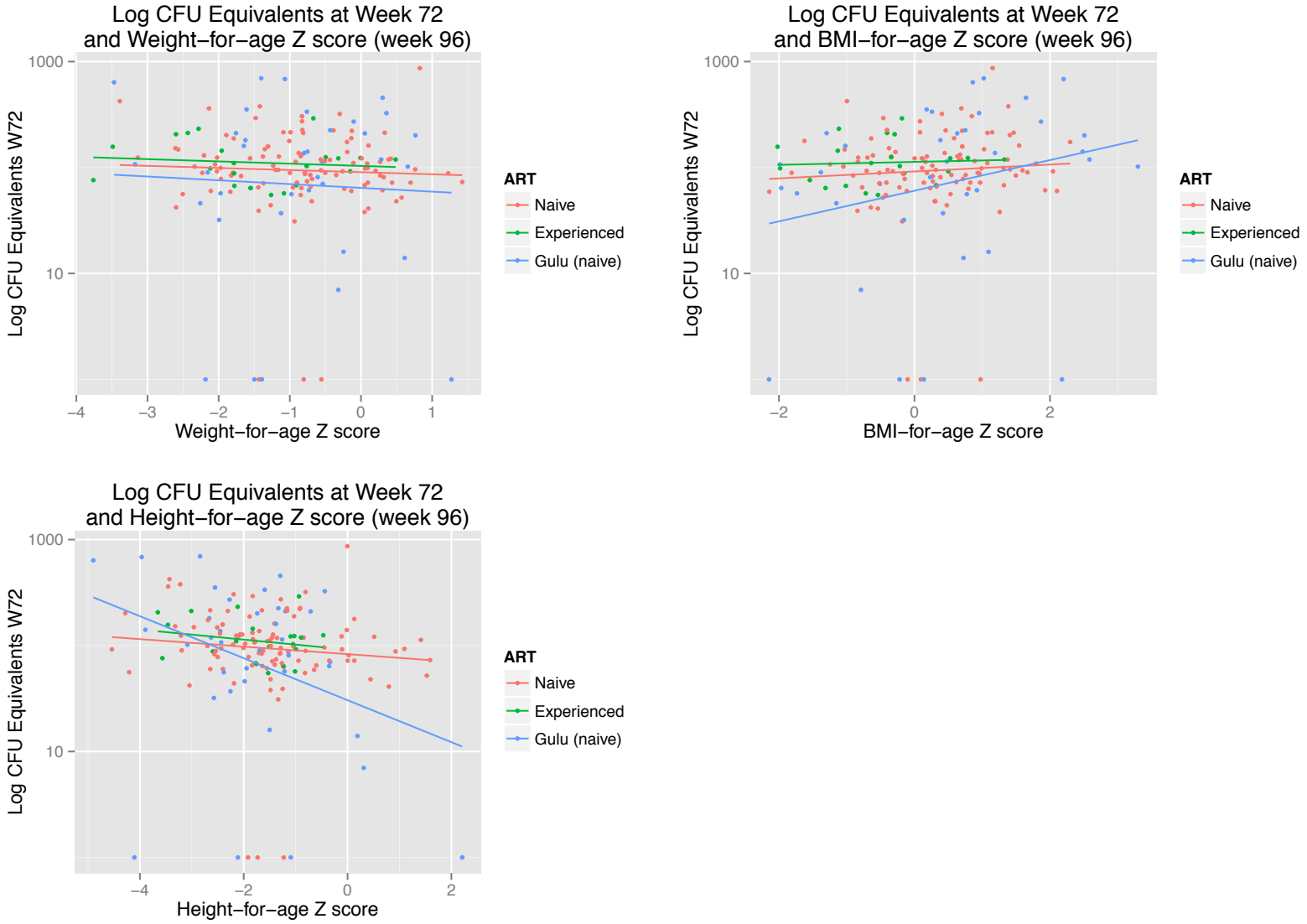


Figure 48 Scatterplot of week 72 bacterial load (CFU Equivalents generated by broad range 16S rDNA PCR using SYBR® Green) compared to week 96 anthropometric measurements

Table 32 Correlation coefficients between bacterial load (CFU Equivalents generated by broad range 16S rDNA PCR using SYBR® Green) and anthropometric measurements

Time point	Spearman Correlation using complete observations		ART Naïve		ART Experienced		Gulu (naïve)		Naïve-control		Experienced-Control	
			Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
Baseline	**CFU equivalents**	Weight-for-age Z score	-0.07	0.44	-0.04	0.87	0.2	0.16	-0.12	0.62	-0.03	0.75
	CFU equivalents	Height-for-age Z score	-0.06	0.54	0.17	0.44	-0.02	0.86	0.03	0.8	-0.08	0.73
	CFU equivalents	BMI-for-age Z score	-0.09	0.33	-0.12	0.61	**0.28**	**0.05**	-0.09	0.43	-0.25	0.29
Week 72 (CFU equivalents)/ Week 96 (anthropometry)	**CFU equivalents**	Weight-for-age Z score	-0.12	0.24	-0.15	0.52	0.04	0.81				
	CFU equivalents	Height-for-age Z score	**-0.21**	**0.03**	-0.11	0.64	-0.29	0.08				
	CFU equivalents	BMI-for-age Z score	0.11	0.28	0.10	0.68	**0.36**	**0.03**				

15.4.4 Comparison of *S.aureus* and *Enterobacteriaceae* qPCRs with Anthropometric Measurements

Baseline and week 72 qPCR results for the *S.aureus* and *Enterobacteriaceae* (coded from negative to strong positive) were compared with baseline and week 96 anthropometry results using the schema in Table 10.

For *S. aureus,* there were no significant associations in any of the three HIV-infected groups, except for a negative association between the Gulu (rural, ART-naïve) group week 72 *S.aureus* results and the week 96 height-for-age z-scores (rho=-0.34, p=0.04) (Table 33). Correlation coefficients could not be calculated for the control groups due to insufficient finite observations.

Considering the *Enterobacteriaceae* assays, there was a significant negative relationship between the height-for-age z-scores of the ART-experienced group at baseline and the level of *Enterobacteriaceae* (rho=-0.46, p=0.03)(Table 34). At week 72, there was a significant negative relationship between the BMI-for-age z-score of the ART-naïve group and the *Enterobacteriaceae* levels (rho=-0.20, p=0.04). Once more correlation coefficients could not be calculated for the control groups due to insufficient finite observations.

Comparing the results of the pellet samples with the anthropometric measurements, for *Enterobacteriaceae,* no relationship was found for the ART-naïve (rho=-0.01, p=0.9), or ART-experienced (rho=0.31, p=0.16). There was similarly no significant relationship between the *S.aureus* results and anthropometric measurements for either ART-naïve (rho=0.03, p=0.71) or ART-experienced (rho=-0.22, p=0.32).

Table 33 Correlation coefficients between *S.aureus* coded negative to strong positive and anthropometric measurements at Baseline and week 72.

Time point	Spearman Correlation using complete observations		ART Naïve		ART Experienced		Gulu (naïve)	
			Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
Baseline	**qPCR result categorised from negative to strong positive**	Weight-for-age Z score	-0.09	0.33	0.29	0.19	-0.09	0.55
		Height-for-age Z score	-0.05	0.57	0.17	0.45	-0.09	0.51
		BMI-for-age Z score	-0.14	0.13	0.46	0.03	0.03	0.83
Week 72 (qPCR results)/ Week 96 anthropometry		Weight-for-age Z score	-0.04	0.65	0.18	0.42	-0.06	0.71
		Height-for-age Z score	-0.02	0.85	0.3	0.19	**-0.34**	**0.04**
		BMI-for-age Z score	0.04	0.7	0.15	0.52	0.29	0.08

Table 34 Correlation coefficients of *Enterobacteriaceae* coded negative to strong positive versus anthropometry

Time point	Spearman Correlation using complete observations		ART Naïve		ART Experienced		Gulu (naïve)	
			Correlation coefficient	P value	Correlation coefficient	P value	Correlation coefficient	P value
Baseline	**qPCR result categorised from negative to strong positive**	Weight-for-age Z score	-0.03	0.75	-0.35	0.12	0.05	0.75
		Height-for-age Z score	-0.13	0.17	**-0.36**	**0.03**	-0.12	0.39
		BMI-for-age Z score	0.14	0.14	-0.13	0.56	0.21	0.15
Week 72 (qPCR results)/ Week 96 anthropometry		Weight-for-age Z score	-0.04	0.65	0.18	0.4	-0.05	0.78
		Height-for-age Z score	0.11	0.25	0.11	0.63	-0.05	0.77
		BMI-for-age Z score	**-0.20**	**0.04**	0.13	0.58	-0.07	0.66

15.5 Discussion

From these results, it appears that I-FABP increased significantly with time in both ART groups, but to a greater extent and in a more sustained fashion (to week 72) in the ART-experienced group. This increase over time brought the levels in both HIV infected groups from JCRC (urban) up to levels seen in uninfected controls in the same community. Interestingly, the Gulu (ART-naïve) group from a rural setting had the highest levels of all groups, which remained consistently high over time. These results appear to run counter to the *a priori* hypothesis that HIV related gut damage would cause higher levels of I-FABP in the ART-naïve group compared with ART-experienced and HIV uninfected children as has been seen in HIV-infected adults in developed settings[63,83,84]. In Ugandan HIV-infected adults, I-FABP was actually lower in rapid progressors than slow progressors, with no significant difference between infected adults and controls. Indeed, I-FABP in HIV uninfected adults was similar to that in HIV-infected adults in developed settings.[85] In this Ugandan study, the median level in HIV-uninfected adults was 330 pg/mL (IQR 68-2861), and in HIV-infected adults was 160 pg/mL (100-1592), tending to be higher than the urban groups in this study (baseline median in ART naïve children (urban) 74 pg/mL, IQR 40-125, in ART-experienced 72 pg/mL IQR 53-111) although not dissimilar to the results seen in the Gulu (rural) group (median at baseline 249 pg/mL, IQR 189-296). I-FABP levels seen in the HIV-uninfected control groups were 140 pg/mL (IQR 75-222) and 160 pg/mL (IQR 65-305) in age-matched naïve and age-matched experienced respectively.

In terms of further external comparison with studies from developed settings, the levels detected in 27 HIV-infected adults in a French study (14 of whom had initiated ART, 13 not receiving ART) were higher 6 months after HIV infection than health donors, with levels ranging between 0-2000 pg/mL, with the majority having levels >500pg/mL, higher than those recorded in any group in this study[63]. Median levels of I-FABP found in Italian HIV-infected adults on ART were 721 pg/mL and 868 pg/mL in those responsive and unresponsive to ART respectively, again, higher than seen in this study[86]. The levels seen in HIV-infected adults in SMART study based in the

U.S.A. were more comparable to those seen in this study with median varying from 70-175pg/ml depending on group.

The lower overall levels of I-FABP in comparison with those seen in developing settings, raises the question as to whether the high prevalence of gastrointestinal infections (in particular parasitic diseases) in resource-limited settings leads to a different paradigm of enterocyte turnover. It has been suggested that constant remodelling of the gut by enterocyte turnover may protect the gut against persistent parasitic infection[87]. The results presented here also support this alternative paradigm, that a higher I-FABP may be a marker of a healthier gut in African children as seen in HIV-uninfected controls. Considering that the children from Gulu (rural) had the highest levels, this might indicate a higher parasite exposure than the children from JCRC (urban) and so a greater enterocyte turnover level. The children from Gulu were of similar ages with similar baseline anthropometry measurements and gains in growth after ART commencement to the ART naïve group from JCRC (Figure 22, Figure 23), and all children received deworming on ART initiation, indicating that the relationship between I-FABP, nutritional status and gut parasites in HIV-infected children in this setting is likely to be complex.

In comparing I-FABP to bacterial load (CFU equivalents generated by broad range 16S rDNA PCR), there was borderline significance of a negative correlation between the two markers at week 12 in the ART-experienced group (Spearman correlation coefficient -0.47, p=0.04). This should be interpreted with caution as the sample size was small (n=19) and multiple comparisons have been made so there is a possibility that at there will be some statistically significant associations seen due to chance.

The anthropometric markers over time were compared to the results of the molecular assays. Although some significant associations were seen, these were limited to one anthropometric marker and one time point, with the exception of a positive relationship between CFU equivalents by broad range 16S rDNA PCR and the BMI-for-age z-score at both time points. As has been previously discussed, these

molecular assays were working at the limits of detection, and sample size for ART-experienced groups and even the Gulu group at the later time point was small. The possibility that the associations seen were spurious must be considered. This is particularly to be considered in view of the number of tests being carried out. As highlighted above, I would anticipate 7-8 tests to be positive at $p<0.05$ due to chance alone. No significant relationships were identified at a level of $p<0.01$, so it is possible that the relationships identified between anthropometric markers and molecular assays were due to chance.

In terms of limitations of the I-FABP, as with the other assays, the sample size for the ART-experienced group was small. Additionally there may have been some variation within the assay itself that could have been compensated for by performing the assay in duplicate. However the volumes of samples available were small and limited repetition. In comparison to the molecular assays, this ELISA based test is likely to be less vulnerable to contamination and variability. As with the molecular assays, samples from the Gulu group were significantly more likely to be missing than the other two groups, which renders the Gulu results more susceptible to bias. As the characteristics of the children for whom data were missing were similar to those without missing data, this bias may be minimal but should be taken into account.

In terms of the anthropometric markers, again, the Gulu group was more likely to have missing data, incorporating a potential bias into the results. Furthermore, there are ongoing challenges about the most representative measures and reference groups to use when monitoring the growth and nutritional status of HIV-infected children in these settings[88,89]. Z-scores are the internationally accepted standard (with provisos as regards the reference population as previously highlighted), but each measure has its own limitations and should be interpreted together with other measures if available. Furthermore, trajectories are more important than single measurements, and although two time points are available here, further monitoring of both I-FABP and the growth of these children over time would be helpful to clarify the relationship between growth and I-FABP in this setting.

Overall, it appears that there were significant differences between the ART groups over time, with both urban groups seeing an increase in I-FABP over time. This might be explained by a different underlying mechanism of change in this population of HIV-infected and uninfected African children compared with HIV-infected adults in developed settings. Examining the relationship between anthropometric markers, disease progression and I-FABP over time merits further specific investigation.

16 Chapter 16 Clinical Events in HIV-infected Children during the Course of the Study in Relation to Molecular Markers of Microbial Translocation

16.1 Background

HIV-infected children are known to be at high risk of bacterial infections, despite ART[90]. There might be a relationship between microbial translocation and these infections. In order to investigate this potential relationship, and to investigate the potential relevance clinically of positive molecular assay results obtained in this study, the relationship between clinical events and assay results were compared.

16.2 Study Population

See section 9.1

16.3 Methods

See section 9.9. Where two events were recorded on the same day, the primary event name only was included (i.e. if the patient had malaria and also had anaemia recorded as event 2, only malaria was recorded). For analysis some types of events were combined. These were: pyogenic meningitis and meningitis- other; any kind of tuberculosis; any kind of pneumonia (pneumonia- other bacterial, pneumonia – pneumococcal, pneumonia no organism identified); rash erythematous and rash maculopapular, upper respiratory tract infection – not sinusitis or otitis media and other acute upper respiratory symptoms.

16.4 Results

There were 386 events reported over the duration of the trial, including hospital admissions, adverse events as defined by the trial protocol and microbiological investigations sent (Table 35). 29 of these occurred in children who also had either a high qPCR /16S rDNA PCR result or a sample successfully sequenced with NGS within four weeks of the event, 72 occurred in children who had study samples taken within four weeks of the event but assay results were negative, and 287 occurred in children who had no study samples within four weeks of the event. The most frequent events were pneumonia (89 events), presumed sepsis or bacteraemia (64 events), malaria (49 events) and upper respiratory tract infections (31 events).

To take first the comparison between children who had study samples available within four weeks of the clinical event, children with positive assay results were of similar age to those with negative results (median age 1.8 years, IQR 1.2-2.3 versus median 2 years, IQR 1.8-4)(Table 35). The children with positive assay results were all from the naïve group, unlike those with negative assay results of whom 29 (40%) came from the Gulu group and one (2%) from the ART experienced group (Fisher's exact test $p<0.001$). There was no difference in the proportion of children admitted ($p=0.52$), nor in the overall distribution of event types ($p=0.13$) between the groups. Considering those who did not have a study sample available within four weeks of the event, these children were older than those with a positive assay result within four weeks of the event ($p=0.03$) and came from all three HIV-infected groups (Fisher's exact test $p<0.001$). There was no difference in the proportions of children admitted ($p=0.01$), but there was a difference in the distribution of event types ($p<0.001$).

Looking in more detail at the events which occurred within 1 month of a high positive test result or successfully sequenced using NGS, 22 (21 patients) of these had high cell pellet results (>1000 CFU equivalents using 16S rDNA PCR with SYBR® Green), one of which also had a high plasma result using 16S rDNA PCR and a high positive with *S.aureus* PCR (Table 36). Six further events took place within a month of a sample having successfully sequenced with NGS.

For children who had either a positive blood or urine culture at any time point, the molecular test results at all time points are documented in Table 37. There were no cerebro-spinal fluid results recorded within the study population. Positive blood/urine cultures included: *S.aureus, Staphylococcus sp., E. coli, Klebsiella pneumoniae, Candida albicans* and *Candida sp.* The episode of *E.coli* bacteraemia was at week 60. The patient had low levels of *Enterobacteriaceae* (9 CFU equivalents by qPCR in baseline pellet sample, and 4 CFU equivalents at week 12), but the week 72 sample was negative. The episode with a positive blood culture for *S.aureus,* was six days after a negative plasma PCR for *S.aureus* and a pellet sample that had a CFU equivalent result of 80 (low positive), and the *Staphylococcus sp.* bacteraemia

episode was at week 85, preceded at week 72 by negative qPCRs. The *Klebsiella pneumoniae* episode was at week 88, and preceded by negative qPCRs apart from one low level positive *S. aureus* PCR in a baseline pellet sample. There was one event of chronic suppurative otitis media with an ear swab result positive for *S. aureus* with a positive qPCR result for *S.aureus* on the same day. However, a further positive swab result for *S. aureus* from a septic ulcer was not accompanied with a positive qPCR result, although there were staphylococcal OTUs in the NGS result (Table 36).

To compare distribution of OTUs amongst children who had a clinical event versus those who did not, the PCoA plot (using unifrac) was repeated, using clinical event versus no clinical event as a differentiating feature (Figure 49, Figure 50). No clustering of OTUs was seen by clinical event/no event.

Table 35 Comparison of clinical events occurring within one month of a high qPCR/16S rDNA PCR result or successful NGS sequencing

	SYBR/QPCR positive within 4 weeks of event	SYBR/QPCR negative AND event <4 weeks before/after study sample timing	P value event & positive sample versus event & negative study sample	No study sample <4 weeks of Event	P value event and positive sample versus event and no study sample available
Number	29	72		287	
Median age (IQR)	1.8 (1.2-2.3)	2 (1.3-3.7)	0.48	2.9 (1.8-4)	**0.03**
ART group					
Naïve	29	42	**<0.001**	176	**<0.001**
Experienced	0	1		24	
Gulu (naïve)	0	29		87	
Admitted	15 (52%)	32 (46%)	0.52	104 (36%)	0.11
Event Name			0.13*		**<0.001***
Meningitis	0	0		2	
Epilepsy- fits- convulsions	0	0		2	
Toxoplasmosis of the brain	0	2		3	
Acute otitis media	0	4		13	

Chronic otitis media	4	1	11
Upper Respiratory Tract Infection	1	4	27
Tuberculosis	0	1	6
Pulmonary Cryptococcus	0	1	0
Pneumonia	7	25	57
Conjunctivitis	0	1	2
Diarrhoea	2	9	20
Urinary Tract Infection	1	1	9
Candida Balanitis	0	0	1
Herpes Zoster (Varicella Zoster) - cutaneous	0	1	0
Rash - erythematous or maculopapular	1	0	6
Tinea - athletes foot - fungal infection of foot or skin	0	0	3
Cellulitis	0	0	2
Impetigo	0	1	1
Folliculitis - furuncles - carbuncles	0	0	1
Lymphadenopathy	0	1	0
Anaemia with clinical symptoms	0	1	2
Anaemia with no clinical symptoms	1	0	0
Kaposi's sarcoma lymph nodes	0	2	0
Malnutrition	0	1	1
Presumed septicaemia/bacteremia - not investigated/ no organism	6	10	54
Malaria	3	4	45
Measles	0	0	3

Acute febrile episode - undiagnosed	2	2	8
Unknown	0	0	1
Not available	0	0	7
	***Comparing whole distribution of events for group rather than individual types of event**		

Table 36 Description of clinical events occurring within one month of a positive test molecular test result (all children were ART naïve and from the JCRC cohort)

Age	week	Sample Date	Pellet Bacterial load (CFU Equivalent)	Plasma Bacterial load (CFU Equivalent)	NGS success	QPCR positive	Clinical event within 4 weeks of positive test	Admitted 0=no, 1=yes	Episode Date
3.5	-1	24-Mar-11	1646		Yes		Presumed Sepsis, blood culture negative	No	13.4.11
1.8	0	07-Jul-11	2294		Yes		Pneumonia, diagnosed 1 month later with Kaposi Sarcoma. Blood culture: *S.aureus.* Ear swab mixed growth. See table	Yes	13.7.11
0.8	-1	27-Sep-11	2874		Yes		Chronic otitis media, not admitted, no investigation recorded	No	8.10.11
3.5	-1	05-Oct-11	1268		No		Upper respiratory tract infection	No	24.11.11
1.8	-2	03-Oct-11	4868		Yes		Pneumonia, No investigation recorded	Yes	27.10.11
3.1	-1	24-Nov-11	1070		Yes		anaemia	No	12.12.11
3	-2	30-Mar-11	2027		Yes	*S. aureus*	Chronic suppurative otitis media, *S. aureus* on ear swab	No	30.3.11
2.3	-1	05-May-11	1275	5431	Yes		Undiagnosed fever	No	12.6.11
0.7	-2	15-Mar-11	1250		Yes		Diarrhoea & possible urinary tract infection	No	4.4.11 & 1.4.11
1.1	-1	04-Nov-11	4095		Yes		Pneumonia. No investigation recorded	Yes	30.11.11
4.4	0	09-Nov-11	1012		Yes		*Plasmodium falciparum*	No	28.11.11

1.1	-1	13-Sep-11	1408		No	Presumed sepsis & anaemia, no investigation recorded	Yes	13.10.11
3.3	-1	17-Jun-11	2251		Yes	Chronic suppurative otitis media, ear swab *Proteus mirabilis,*	No	24.6.11
0.8	-4	06-Jun-11	1830		No	Upper respiratory tract infection. No investigation recorded	No	1.8.11
2.9	-1	14-Oct-11	1845		Yes	Bacteraemia, No investigation recorded	No	2.11.11
1.5	-2	28-Oct-11	5996		Yes	Pneumonia	Yes	25.11.11
2.1	-1	04-Nov-11	3995		No	Presumed sepsis	No	15.2.12
3.1	-1	07-Sep-11	1468		No	Presumed sepsis	No	26.10.11
1.5	-5	30-Sep-11	1619		No	Pneumonia. No investigation recorded	Yes	17.11.11
1.8	-5	10-May-11	1759		No	Presumed sepsis	No	26.6.11
4.3	-1	19-Sep-11	6011	6065	No	Chronic suppurative otitis media. No investigation recorded	No	8.11.11
1.2	12	25-May-11		82	Yes	Acute Diarrhoea. No investigation recorded	Yes	11.4.11
1.2	-3	8-Apr-11	1272	0	Yes	Plasmodium falciparum malaria	Yes	16.5.11
4.4	12	2-May-11			Yes	Septic ulcer, *Staphylococcus aureus* on swab	No	2.5.11
1.6	-2	30-Jun-	37		Yes	Pneumonia	Yes	1.8.11

		11						
3.6	-2	22-Jun-11	487	Yes	Pneumonia		Yes	8.7.11
4.7	-2	14-Sep-11		Yes	Plasmodium falciparum malaria		Yes	13.10.11

Table 37 Assay results for patients with positive blood/urine culture

Age (yr)	Naïve/ Experienced	Trial Week	Blood/ Urine	Organism	Presumed Diagnosis	Admitted	Test time point	qPCR CFU equivalents*	16S CFU equivalent	I-FABP pg/mL	NGS successful
1.8	Naïve	1	Blood	*Staphycocccus aureus*	Pneumonia, Kaposi sarcoma	Yes	baseline pellet	SA: 80	2290		Yes
							baseline plasma			<20	
							W12		200	80	
							W72		272	58	
1.1	Naïve	85	Blood	*Staphylococcus sp* (coagulase negative)	Sepsis/ Urinary tract infection	Yes	baseline pellet	SA: 20, Ent 6.6	91		Yes
							baseline plasma		55	368	
							W12		187	132	
							W72		121	98	
2.7	Naïve	60	Blood	*Escherichia coli*	Presumed Sepsis	Yes	baseline pellet	Ent: 9, Tuf: 4	110		Yes
							baseline plasma		208	31	
							W12	Ent: 3	60	95	
							W72		123	134	
10.7	Experienced	88	Urine	*Klebsiella pneumoniae*	Pneumonia/ Urinary tract infection	No	baseline pellet	SA: 8.3	2980		Yes
							baseline plasma		185	<20	
							W12			<20	
							W72		97	215	
1.3	Naïve	2.3	Urine	*Candida albicans*	Pneumonia/ Urinary tract infection	Yes	baseline pellet		20		Yes
							baseline plasma	Bif: 76	165	162	
							W12	Bif: 6	220	61	
							W72		10	153	
1.8	Naïve	32	Urine	*Candida* sp.	Pneumonia/ Diarrhoea/ presumed	Yes	baseline pellet	SA: 96, Ent: <	2210		Yes
							baseline plasma		<5	64	

	sepsis	W12	108	137
		W72	71	62

*Ent: *Enterobacteriaceae,* SA: *Staphylococcus aureus*, Tuf: *Staphylococcus spp.,* Bif: *Bifidobacterium*

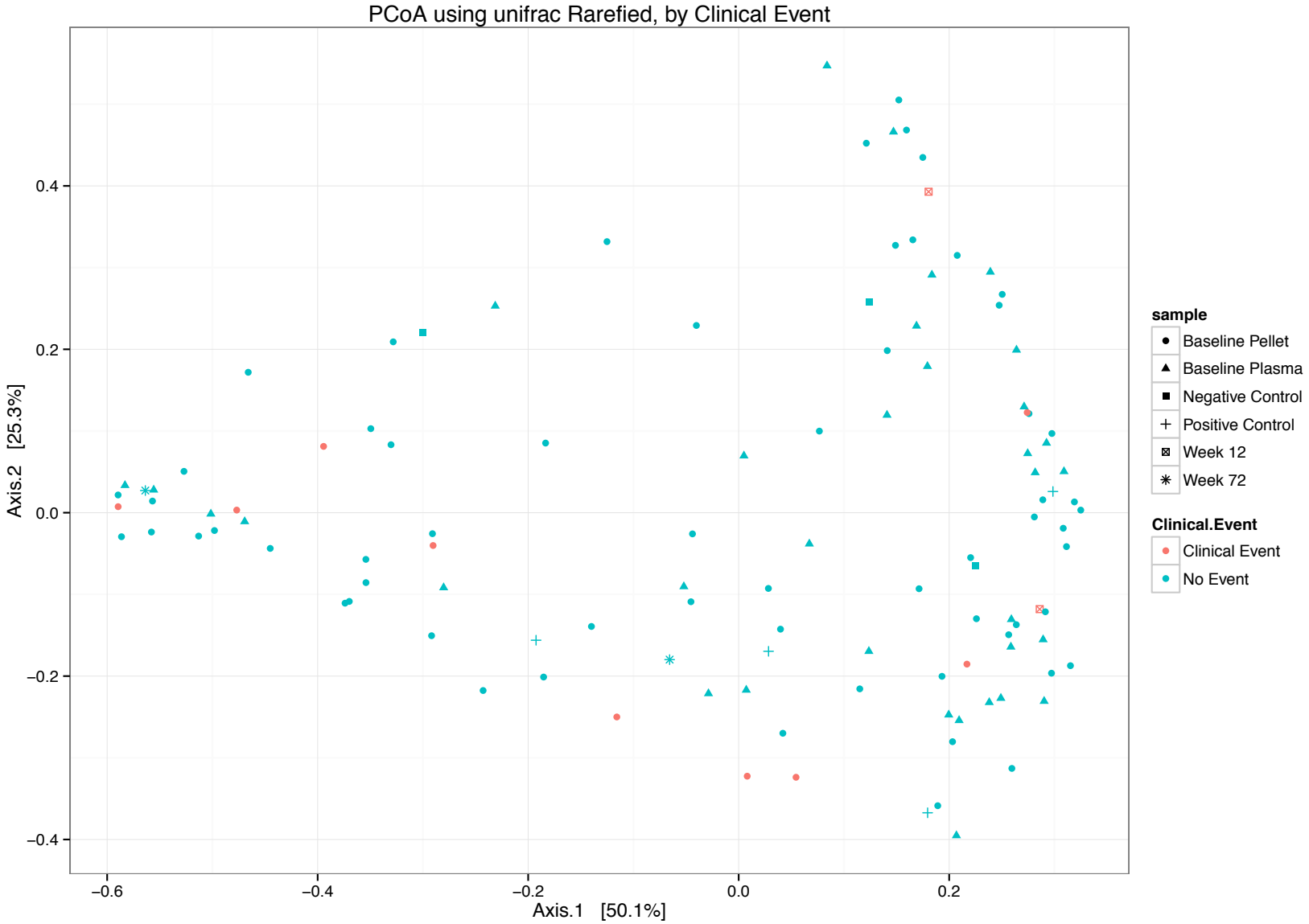


Figure 49 PCoA of Rarefied Samples, by Clinical Event/No Event using unifrac showing no clustering according to whether or not a clinical event occurred.

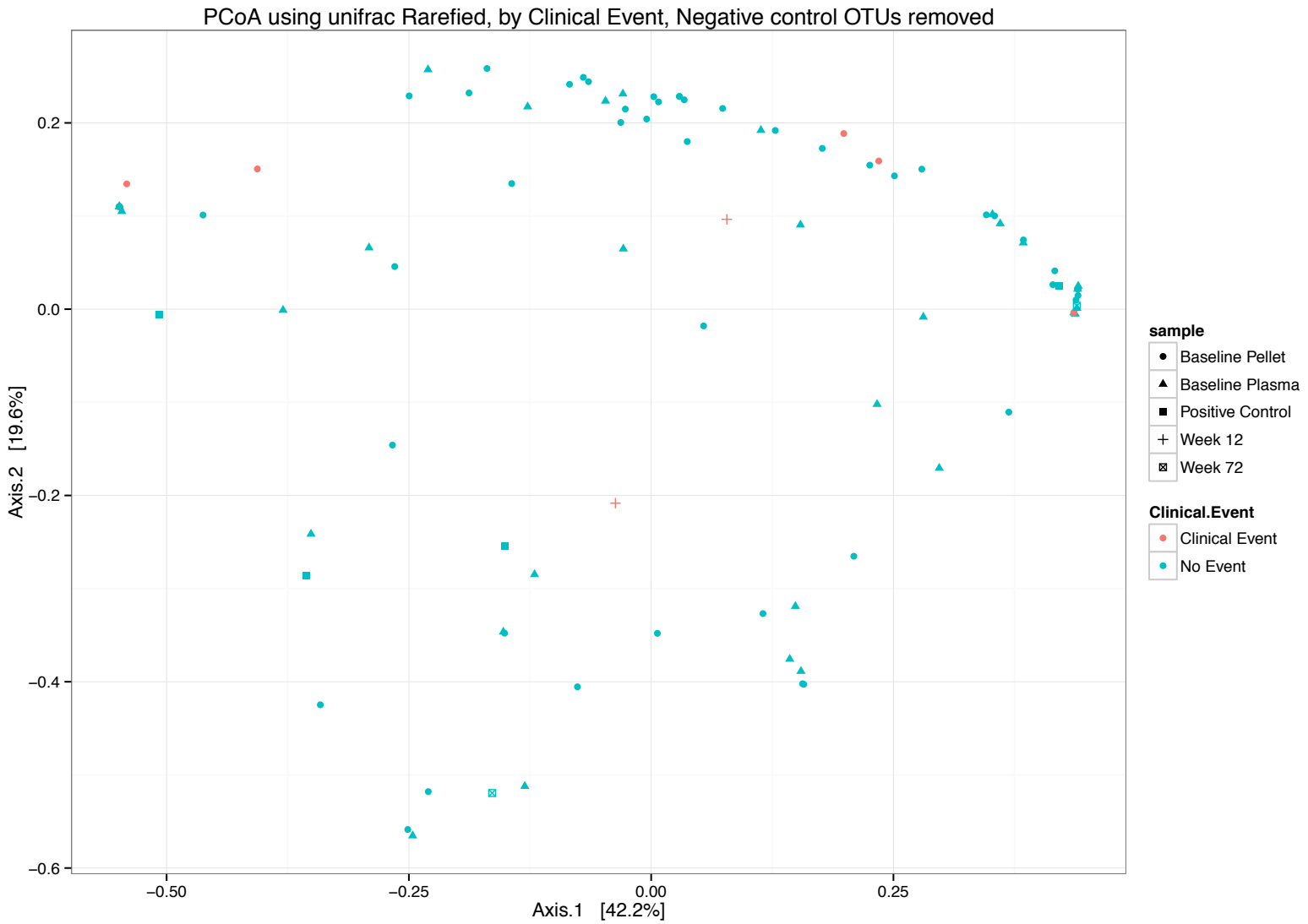


Figure 50 PCoA of rarefied samples (unifrac), negative control OTUs removed by presence/absence of clinical event, showing no clustering by whether or not a clinical event occurred

16.5 Discussion

This section documents the type and frequency of clinical events experienced by children enrolled in the cohort: namely adverse events as defined by the trial protocol, hospital admissions and any positive microbiological test results, and compares these events with molecular test results. First, the frequency and type of clinical event was compared by whether or not the child had also had a high positive qPCR or broad range PCR result, or a sample that successfully sequenced using NGS within a month of the event. There did not appear to be a significant difference between types or frequency of events between the two groups for whom study samples were available within four weeks of the event, although the children who experienced events and had a high test result were all from the JCRC ART naïve group (Table 35). However, when considering this result, it should be seen that the positive test results were overwhelmingly from baseline pellet samples which were not available from the Gulu children, or at later time points, and the ART naïve group was larger than the ART experienced group (and also younger). This indicates that pellet sample availability is likely acting as a confounding factor, as the pellet samples tended to be more likely to have a high result across all molecular assays than the plasma samples (Sections 12.4.2, 13.4.1 and 14.4).

Confounding also applies to consideration of the differences seen between children whom had an event and had no study sample available within four weeks of the event, in that they also were more likely to come from the ART experienced or Gulu groups (neither of which had pellet samples) than those who had an event and a positive assay result. The group with a clinical event but no study sample available were also slightly older, and had a different distribution of event types ($p<0.001$) to

those with a positive assay result. However, commonly occurring events such as malaria, pneumonia and diarrhoea were seen in both groups and admission rates were similar (p=0.11). In view of the similarity between the groups for whom samples were available, it is unlikely that a biologically important difference has been missed in comparing those experiencing clinical events with and without positive assay results.

The assay results from children who had positive blood or urine cultures at any time point were then examined to look for any relationship between the organisms found and molecular test results (Table 37). There was no apparent close relationship between molecular results and blood/urine cultures, though this comparison was limited by sample availability: i.e. samples were only available at fixed time points that in several cases were remote from the event itself.

A limitation of this comparison is the clinical data availability. The diagnoses were taken verbatim from codes documented in the trial documentation but may have been subject to variability and diagnostic uncertainty. For example, many events were coded as “Pneumonia- pneumococcal” with no microbiological evidence that pneumococcal pneumonia had been confirmed. By combining all possible codes for pneumonia (encompassing “pneumonia- other bacterial”, “pneumonia – pneumococcal” and “pneumonia no organism identified”) I attempted to minimise this uncertainty, but a degree of variability is likely to remain. Similarly, by looking specifically at positive culture results from sterile sites (blood and urine) I used the most concrete microbiological evidence available in the comparison rather than, for example, using urinalysis results that are subject to a higher degree of variability. Positive microbiological swab results from other sites were only included if they coincided with an adverse event (as defined in the trial protocol) or an admission, in order to prioritise clinically significant events for the comparison.

This comparison was intended as broad attempt to compare clinical events with molecular assay results, looking for high-level trends or indications that there might

be a relationship. However, although sample availability (in particular only having pellet samples in the JCRC group and only at baseline) is a major limitation, the relationship between assay results and clinical events appears to be small.

17 Chapter 17 Discussion

The aims of this thesis were:

1. To assess the level of microbial translocation defined as identification of bacterial DNA, occurring in HIV-infected African children by comparing HIV-infected ART naïve children with HIV-uninfected African controls
2. To determine the effect of ART on microbial translocation by comparing levels detected in ART-naïve and ART-experienced children, and changing levels over time on ART.
3. To determine the impact of environment on microbial translocation by comparing levels in ART-naïve HIV infected children in urban versus rural settings at baseline and over time on ART.
4. To investigate the relationship between microbial translocation, gut damage and clinical outcome in HIV infected children in Uganda.

In my final discussion I will examine whether we have found evidence of microbial translocation in HIV infected children, if the methods developed are reliable and in the light of my results, what work should be undertaken in the future.

17.1 Does Biologically Significant Microbial Translocation Occur in Children with HIV

At the time this study was designed, clinical trials were underway with the aim of modifying microbial translocation and hence improving outcome in people with HIV. However, on closer scrutiny of available data, the evidence for microbial translocation as an important driver of immune activation was by no means conclusive. Prior studies, summarised in Table 38, were largely small, cross sectional and based in developed settings, and only four studies included children. Furthermore, there was no consensus on the best method of quantifying microbial translocation. The assay used to quantify LPS is subject to considerable variability and plasma inhibition, even within subjects let alone between studies[91]. Additionally, LPS could be considered a

Table 38 Summary of previous studies investigating microbial translocation in the blood of humans infected with HIV, including methods used for quantification and study results (as regards microbial translocation)

Population	Country	Median age in years (range)	Microbial translocation markers measured	Ref
85 ART experienced children over 48 weeks enrolled in the Pediatric AIDS Clinical Trials Group Protocol P338, adult HIV-uninfected controls	India	Range 2-17 Median not given	16S rDNA PCR, sCD14, LPS. All higher than controls, no change over time. 16S rDNA, CD4 T cell count & immune activation associated, no association with LPS or viral load	Pillakka Kanthikiel *et al.*[61]
54 HIV infected infants in the CHER trial, 34 receiving immediate ART, 20 deferring ART until CD4%<25%, compared with 22 HIV exposed, uninfected infants.	South Africa	0-1 year (enrolled at birth)	LPS, sCD14, LBP. All markers and T cell activation were higher in those deferring versus immediate ART. LPS higher than uninfected controls. No LPS in those >6 months of age. No association between LPS and T cell activation	Papasavvas *et al.*[92]
77 HIV infected children, 32 controls; Cross sectional study 70% of HIV-infected virally suppressed.	Spain	16.2 years (13-18.7)	LPS, 16S rDNA PCR. T cell activation associated with HIV viraemia. No association between LPS or 16S rDNA with immune activation, although levels higher than	Madrid *et al.*[70]

			HIV-uninfected controls.	
33 ART experienced children commencing PI-based therapy over 96 weeks, 14 HIV-uninfected controls.	U.S.A.	HIV-infected: 6.6 years(0.2-17.7 Controls: 1.4 years (0.4-15)	LPS, sCD14, EndoCAb. 10% had elevated LPS, associated with increased monocyte activation, but no relationship with sCD14, T cell activation, CD4 T cell count, thymic output or EndoCAb. Decrease in immune activation seen with immune recovery. Healthy infants < 6 months also had raised LPS.	Wallet *et al.* [93]
205 HIV infected (50 acute infection, 30 "elite controllers" with undetectable virus and no ART, 27 with detectable virus and CD4 >250*, 7 with AIDS and CD4<250. Cross sectional.	U.S.A. and France	Adult (age not given)	LPS, sCD14, LBP, EndoCAb. Higher LPS in HIV infected than controls. "Elite controller" had lower sCD14 and LBP, and higher EndoCAb than those with disease progression.	Brenchley *et al.* [94]
Cross sectional study of mucosal integrity in 19 HIV negative, 15 HIV infected untreated, 15 HIV infected on ART CD4>500, 24 HIV infected on ART CD4<350.	U.S.A.	45 years (34-55)	sCD14 sCD14 linked to gut mucosal apoptosis.	Somsouk *et al.* ([95])
Cross sectional study of intestinal dysbiosis in 18 HIV infected adults	U.S.A.	32 years (22-58)	LPS, sCD14, I-FABP, LTA.	Dillon *et al.*[96]

and 14 HIV uninfected controls.			Plasma LTA significantly associated *Bacteroides* abundance at gut mucosa. PCoA revealed association of microbiome patterns with HIV infection status, LPS, colon and peripheral blood T cell activation but no test for association independent of HIV infection.	
Cross sectional study of intestinal dysbiosis in 80 HIV infected (42 no ART, 40 on ART) and 40 HIV uninfected controls.	Uganda	44 (on ART) (38-49) 29 (no ART) (24-34) 43 (controls) (38-48)	sCD14 sCD14 higher in HIV infected (CD4<200*) than controls or HIV infected (CD4>200). sCD14 significantly associated with HIV viral load. 144 bacterial taxa in stool significantly associated with high sCD14.	Monaco *et al.* [97]
Longitudinal study of intestinal dysbiosis in 13 adults with acute HIV infection over 48 weeks after ART initiation.	U.S.A.	33 (21-55)	LPS, sCD14 Increased *Lactobacillales* associated with lower sCD14, higher CD4 T cell percentage and lower viral loads. No association with LPS or with immune activation.	Perez-Santiago *et al.* [7]
Cross sectional study investigating invariant natural killer T (iNKT) cells, gut dysbiosis and gut barrier	U.S.A.	50 (25-66) (HIV) 33 (23-59)	sCD14, LBP, I-FABP.	Paquin-Proulx *et al.* [98]

function in 23 HIV infected (13 on ART) and 10 HIV uninfected controls		(controls)	I-FABP, sCD14 and LBP higher in HIV infected than controls. Inverse relationship between iNKT cells and microbial translocation markers.	
Cross sectional study of 173 HIV infected adults, longitudinal study of 54 adults (48 weeks) and 15 HIV uninfected controls	U.S.A.	48 (40-54) (HIV) 26 (22-41) (controls)	LPS, 16S rDNA PCR Higher levels in HIV-infected than controls, associated with T cell activation. Negative association of 16S rDNA with CD4 T cell restoration after adjusting for viral load at week 48. Those on ART had lower 16S rDNA.	Jiang *et al.* [59]
Longitudinal study of 44 HIV infected adults commencing ART (48 weeks) and 13 HIV negative controls	Italy	38 (29-52) (HIV) 27 (25-32) (controls)	sCD14, LPS, 16S rDNA PCR with Sanger sequencing sCD14 higher in HIV than controls. No change in LPS or sCD14 over time on ART, or difference between those responsive/unresponsive to therapy. Higher proportion of HIV infected positive for 16S rDNA PCR than controls. Mixed sequencing results with no major changes over time on ART. Minimal difference between responsive/unresponsive to therapy.	Merlini *et al.* [8]
Longitudinal study of 420 HIV infected adults in SMART trial and	U.S.A. (90%), 10%	48 (40-56)(HIV) 35 (controls)	LPS, sCD14, EndoCAb, 16S rDNA PCR, I-FABP.	Sandler *et al.* [60]

67 HIV uninfected controls	from 32 other countries	range not given)	Only sCD14 associated with mortality. I-FABP higher in HIV infected than controls, no difference in 16S rDNA between HIV infected/uninfected.	
Cross sectional study of 60 HIV-infected adults investigating association of endovascular markers with microbial translocation.	Denmark	47 (40-65)	sCD14, LPS sCD14 associated with increased risk of cardiovascular events but not LPS.	Pedersen *et al.* [99]
Longitudinal cohort of 107 HIV infected adults without ART.	Uganda	29 (23-43.3)	sCD14, LPS sCD14 and immune activation markers raised but not LPS. No change in LPS over time without ART.	Redd *et al.*[100]
Cross sectional study of 30 HIV infected adults on ART investigating gut damage and response to ART.	Italy	54 (43-71)	sCD14, LPS, EndoCAb C-reactive protein related to disease progression but markers of microbial translocation not related	Tincati *et al.*[86]
Longitudinal study of early HIV infection in 27 adults	France	Adult (age not given)	16S rDNA PCR, I-FABP Immune activation occurred in absence of	Chevalier *et al.* [63]

			microbial translocation.	
Cross sectional study of adults with HIV, comparing adults with/without ART and with/without response to ART.	Italy	40 (30-69)	16S rDNA PCR with Sanger sequencing, LPS LPS lower in those on ART. LPS associated with T cell activation in those unresponsive to ART. Higher proportion of samples positive with 16S rDNA PCR in ART-naive/unresponsive to ART. *Serratia* spp. and *Rahnella* spp. sequenced.	Marchetti *et al.* [54]
***CD4: CD4 T cell count in cells/µL**				

proxy measure of microbial translocation and as it is a component of Gram negative cell walls means that Gram positive bacteria would not be quantified. Finally, the relationship between LPS, immune activation and disease progression in HIV has not been demonstrated consistently across previous studies (Table 38). The relationship between sCD14, immune activation and disease progression has been more consistently demonstrated, but this is not surprising as sCD14 is itself a measure of monocyte activation, which could be driven by other mechanisms in addition to microbial translocation (section 8.4.5). A raised sCD14 should not be conflated with microbial translocation as it too is a proxy marker. Seven previous studies have used broad range 16S rDNA PCR to quantify microbial translocation, but only two have used sequencing to confirm the results are consistent with microbial translocation as opposed to artefact[8,54]. Indeed, the bacteria identified (*Serratia* spp., *Rahnella* spp., mixtures of *Pseudomonadales* and *Burkholderiales)* could be consistent with contamination as opposed to have translocated from the gut.

This thesis set out to determine whether microbial translocation occurs in HIV infected children in an African setting, and whether, if present, that microbial translocation is of biological importance. Aims 1-3 were addressed using three molecular methods: a panel of qPCRs, a broad range 16S rDNA PCR and next generation sequencing (Sections 12.4, 13.4.1, 14.4) in 140 HIV-infected and 109 HIV uninfected children from an urban environment and 55 HIV-infected, ART naïve children from a rural environment. This was the first time, to our knowledge that NGS had been employed to provide definitive identification of any bacterial DNA identified in the samples. For Aim 4, gut damage was assessed by measuring I-FABP, clinical events occurring in the trial were recorded and compared with results of molecular assays and anthropometric markers were assessed over time in all HIV-infected children, and at a single time point in uninfected children.

The outcomes of investigating aims 1-3 are discussed in chapters 12-14. In general, markers of microbial translocation in all 3 molecular assays were low in all groups and at all time points. *S.aureus* was the most reliably and consistently detected bacterial species, and it is questionable as to whether this DNA is derived from the gut

secondary to microbial translocation or from another source, namely skin contamination. Previous studies have not used sensitive assays such as the *S.aureus* qPCR, so it is difficult to extrapolate to what degree previously demonstrated high levels of 16S rDNA may have in fact been due to skin contamination by *S.aureus* or other skin colonising bacteria as opposed to translocated DNA from the gut. Using NGS, bacteria were identified which could be consistent with microbial translocation from the gut such as *Veillonella spp.* and *Fusobacterium* spp.. However, they were found in very low levels, and in both ART-naïve and experienced children at baseline (Figure 39). It is important to note that I was unable to sequence any samples from HIV-uninfected controls, which may mean that there is a biologically significant difference in the quantities of bacterial DNA between HIV infected/uninfected groups. Indeed, several previous studies have also found no evidence of bacterial DNA using a broad range 16S rDNA PCR (without sequencing) in HIV-uninfected controls[54,59]. However this may have been due to the unavailability of pellet samples for the control groups: and both with the qPCRs and the broad range 16S rDNA PCR I was able to identify low levels of bacterial DNA in the control samples.

Enterobacteriaceae have been implicated in both previous studies to use sequencing methods on the blood of people with HIV[8,54]. Although in this study, *Enterobacteriaceae* were also detected using the qPCR panel and the NGS method, the levels of *Enterobacteriaceae* detected using the qPCR assay were consistent over time in all HIV-infected groups (section 12.4.1), indicating that the results may be due to contamination or of doubtful biological significance. There were differences between ART-naïve and HIV-negative controls at baseline (23% versus 38% positive, p=0.03), but the proportions of positive samples were higher in controls than HIV-infected children. Furthermore, in the context of numerous significance tests, it is possible that these results are spurious, particularly when taken in the context a) of minimal difference between other groups b) understanding the inherent variability of the assay and vulnerability to contamination. There was also minimal association between markers of microbial translocation and clinically significant events in these children over time as discussed in chapter 16.

This is the most comprehensive study to evaluate if microbial translocation is occurring in HIV infected children. To conclude, if translocation is occurring, it is at

very low levels and is unlikely to be of clinical significance. Our data is consistent with previous studies and together they raise doubts about the importance of microbial translocation as a driver of immune activation in people with HIV.

The results for Aim 4 are discussed in Chapters 15-16. Considering the results of the I-FABP assay, the results ran counter to the hypothesis that I-FABP as a marker of gut damage would be higher in ART-naïve HIV-infected children than ART-experience children or in HIV-uninfected controls. Significant increases in I-FABP were seen in both HIV-infected urban groups over time on ART, bringing the levels in line with that seen in HIV-uninfected controls and also that seen in HIV-infected, ART-naïve children from a rural environment (

Figure 44, Figure 45) It appears that in this setting, I-FABP may be a marker of health gut turnover rather than gut damage, but the relationship appears to be complex and further research is needed. An adjusted model of I-FABP and individual predictor variables could contribute to understanding, as could more longitudinal follow-up and a larger sample size. Indeed, I-FABP is only one marker of enteropathy. A panel of other markers that capture permeability, regeneration of the epithelium, inflammation of the gut might provide more information about the extent of gut damage and its relationship to microbial translocation.

17.2 Strengths and Limitations

17.2.1 Assay Limitations

One of the main challenges in investigating microbial translocation is the lack of consensus as to the best methods of quantification. In order to address the thesis aims of definitively answering the question as to whether microbial translocation occurs, I used three molecular approaches. The assays I developed were: a panel of qPCRs, a broad range 16S rDNA PCR (SYBR® Green) and next generation sequencing (NGS) using Illumina® technology. The challenges involved in development of these assays are described in Section 10. It became clear that across all sample types from different populations, levels of bacterial DNA were low, and these assays were working at the limits of detection. Methods were developed accordingly in order to

overcome these challenges, particularly with regards to the NGS. The discovery that the primers were contaminated with *Sphingomonas spp.* DNA (section 10.2.1) and also identification and minimisation of the primer dimer that was interfering with successful sequencing (10.2.2) led to development of a workflow that allowed the bacterial 16S rDNA present in 22% of samples to be amplified to a level that sequencing was possible. However, even for samples that were successfully sequenced, after removal of OTUs that were also present in negative experimental controls, the remaining phylogenetic trees demonstrated only scanty levels of bacteria in both ART naïve and ART experienced children (Section 14.4.1). Although it is possible that these bacteria were truly the result of microbial translocation, at such low levels it is difficult to draw definitive conclusions.

As these assays were working at the limits of detection, it is possible that they were insufficiently sensitive to identify bacteria that had translocated from the gut but were present at very low concentrations. In this study, we did not explore other surrogate markers of microbial translocation such as LPS or sCD14 which might have assisted in identification. However, LPS is a variable assay in itself, with widely divergent results demonstrated in different studies, and no overall consensus as to what should constitute a strongly positive result[60,61,100]. It is also vulnerable to the effects of plasma inhibition depending on the specific patient sample[91]. Indeed, LPS has not been consistently found to be related to disease progression in longitudinal studies of HIV[69,70,100]. Although the evidence of association of sCD14 is stronger than LPS, sCD14 is a marker of monocyte activation not specifically microbial translocation[60]. It could be raised directly by HIV, or by other unquantified environmental factors[101,102].

Furthermore, all three molecular assays gave similar results of very low levels of bacterial DNA across all groups. This strengthens the interpretation that if microbial translocation is occurring, it is happening at such low levels in all groups that the microbial translocation itself may be of little biological importance. Although each assay had its own limitation in terms of sensitivity and specificity, using all three together increased the likelihood of identifying translocating bacteria. For example, although it was not possible to design a qPCR assays that specifically identified *Clostridiales,* bacteria from this family were identified by NGS, namely *Veilonellaceae*

and *Clostridiaceae.* Similarly, *Lactobacillaceae* were identified using NGS and these were taxa that were potentially from a different family to that targeted by the qPCR assay. Perhaps, as concluded by other studies, it may be an aberrant host response to ubiquitous low levels of microbial translocation that is problematic, rather than the microbial translocation itself[60]. It might be similar to stunting, in that although enteropathy is widespread, only one third of infants experience stunting as a result; and that stunting is significantly associated with raised inflammatory makers[103,104]. Potentially, for some individuals microbial translocation is consequential and for others it has no impact.

17.2.2 Limitations in Availability of Samples

Across all three molecular assays, the pellet samples appeared to have high bacterial loads, likely for the reasons discussed in section 14.5 about the potential advantage of greater quantities of host DNA acting as carrier DNA and enhancing the extraction efficiency of the target DNA. However, pellet samples were only available at baseline for the HIV-infected urban groups, so comparisons could not be made either with HIV-negative children, or over time, or between urban and rural areas. In particular this may have affected NGS results, as while 75% of pellet samples sequenced successfully, only 11% of plasma samples did. Therefore baseline samples from the HIV-infected group in the urban setting were likely to be over-represented in the NGS data.

Additionally, the sample size for the ART-experienced group and their age-matched controls was small. Therefore fewer data were available for comparison either between these two groups or with the ART-naïve groups.

17.2.3 Cotrimoxazole

All participants enrolled in the CHAPAS-3 trial received co-trimoxazole prophylaxis from baseline onwards. This might have impacted on results. However, there were minimal differences seen between ART-naïve and ART-experienced at baseline, when the ART-experienced children would have been receiving cotrimoxazole. Elsewhere, in one small adult study in a developed setting, initiating ART with cotrimoxazole prophylaxis appeared to decrease sCD14 and LPS binding protein more than ART

alone[105]. However, in adults in Uganda, it appeared that cotrimoxazole had little impact on the diversity of the microbiome, although the control population were HIV negative as nearly all the HIV infected participants were receiving prophylaxis[97].

In summary, in view of little difference seen between ART-naïve and ART-experienced groups at baseline, it appears unlikely that cotrimoxazole had a significant impact in this study.

17.3 Future Research

This thesis demonstrates the challenges faced when applying molecular methodologies to the investigation of microbial translocation in HIV. An NGS method optimal for sequencing bacterial 16S rDNA in low-biomass samples was developed. It is unlikely that any further gains in sensitivity could be achieved using this methodology. Pellet samples contained the most bacterial DNA and appear to be a superior sample type with assay than plasma for detection of very low levels of bacterial DNA. This method could be used to investigate microbial translocation in other populations, not only in HIV-infected children in other resource-limited settings, but also in HIV-infected individuals in developed settings to explore a relationship with other surrogate markers of microbial translocation and immune activation. It could also be used in other populations where microbial translocation is thought to be of significance such as preterm neonates, surgical patients, patients with inflammatory bowel disease and immunocompromised patients in general[106-109].

Furthermore, with additional quality control, it might be feasible to adapt the method for use clinically, enabling batching of large numbers of clinical samples as opposed to 16S rDNA PCR with Sanger sequencing as is used in clinical settings currently[110,111].

In terms of microbial translocation as an important cause of poor clinical outcome in HIV-infected children in Uganda, on the basis of the lack of definitive evidence provided by this large study, with HIV-negative controls, in conjunction with the lack

of efficacy of plot studies and clinical trials targeted to modify microbial translocation and improve outcome in HIV-infected patients (section 8.4.4), it is difficult to justify further clinical trials at this point in time. Further studies using these techniques may yield different results in different populations, such as very young infants for example who may have a more vulnerable gut barrier.

In terms of immune activation, exploring link between microbial translocation and immune activation in this setting is beyond the scope of this thesis, but future work including cluster analysis of the link between markers of immune activation and microbial translocation results in this study would be useful. Markers of immune activation, inflammation and disordered thrombogenesis decreased over time on ART in the wider CHAPAS-3 cohort[112]. Given the absence of any change over time or between groups in microbial translocation markers in this cohort, a direct relationship between immune activation and microbial translocation appears unlikely. However, this question should be addressed directly in future analysis.

Considering immune activation more broadly, trials targeted to reduce immune activation and improve outcomes for those with HIV have not yet yielded convincing results (section 8.4.4). However, there is merit in exploring other mechanisms of immune activation such as chronic infection with CMV for example. REALITY, a large randomised control of severely immunocompromised HIV-infected adults and older children aiming to reduce early mortality on ART demonstrated a 25% reduction in mortality with a package of enhanced OI prophylaxis, indicating there is considerable benefit to be achieved in reducing the pathogen load for the severely immunocompromised initiating ART.[113]

The results from the I-FABP assay are intriguing and merit further investigation. It may be that in this setting there is a different underlying mechanism of gut turnover. For these children, in the context of high background prevalence of intestinal parasites, it may be that high turnover and high I-FABP is the marker of a healthy gut, and that gains are made over time on ART[87]. There also appeared to be a higher level of I-FABP in the rural group with no significant change over time. We did not have

access to data on the prevalence of parasitic infections in these groups, although all would have received deworming as part of the CHAPAS-3 protocol. An avenue for future research could interrogate further how environment and parasitic infection impacts on gut turnover in this setting.

To conclude, it appears that microbial translocation as detected by these 3 molecular assays is unlikely to be a significant driver of poor outcome in HIV-infected children in this setting. There are technical issues with the sensitivity of the assays used, so future studies may yield more information. However, pragmatically, it appears that there may be more cost-effective ways to intervene to improve outcome for HIV-infected individuals over and above commencing ART, particularly for the severely immunosuppressed such as the enhanced OI prophylaxis package used in the REALITY trial[113]. On a larger scale, with the new WHO guidelines advocating ART for all at the point of diagnosis, the resources of both ministries of health and research institutes are going to be stretched to define innovative and affordable ways of making this ambitious goal a reality. Faced with the dramatic clinical benefits of early ART in improving mortality, it is difficult to see that the gains achieved by targeting microbial translocation will be anything other than marginal.

18 References

References

1. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; **7**(5): 335-6.

2. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* 2006; **72**(7): 5069-72.

3. Crane RJ, Jones KD, Berkley JA. Environmental enteric dysfunction: an overview. *Food and nutrition bulletin* 2015; **36**(1 Suppl): S76-87.

4. Prendergast A, Kelly P. Enteropathies in the developing world: neglected effects on global health. *The American journal of tropical medicine and hygiene* 2012; **86**(5): 756-63.

5. Ellis CL, Ma ZM, Mann SK, et al. Molecular Characterization of Stool Microbiota in HIV-Infected Subjects by Panbacterial and Order-Level 16S Ribosomal DNA (rDNA) Quantification and Correlations with Immune Activation. *J Acquir Immune Defic Syndr* 2011.

6. Dinh DM, Volpe GE, Duffalo C, et al. The Intestinal Microbiota, Microbial Translocation and Systemic Inflammation in Chronic HIV Infection. *J Infect Dis* 2014.

7. Perez-Santiago J, Gianella S, Massanella M, et al. Gut Lactobacillales are associated with higher CD4 and less microbial translocation during HIV infection. *AIDS* 2013; **27**(12): 1921-31.

8. Merlini E, Bai F, Bellistri GM, Tincati C, d'Arminio Monforte A, Marchetti G. Evidence for Polymicrobic Flora Translocating in Peripheral Blood of HIV-Infected Patients with Poor Immune Response to Antiretroviral Therapy. *PLoS One* 2011; **6**(4): e18580.

9. Fitzgerald F, Harris K, Doyle R, Alber D, Klein N. Evidence that Microbial translocation occurs in HIV-infected children in the United Kingdom. *AIDS research and human retroviruses* 2013.

10. Gori A, Tincati C, Rizzardini G, et al. Early impairment of gut function and gut flora supporting a role for alteration of gastrointestinal mucosa in human immunodeficiency virus pathogenesis. *J Clin Microbiol* 2008; **46**(2): 757-8.

11. Smith MI, Yatsunenko T, Manary MJ, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* 2013; **339**(6119): 548-54.

12. Suau A, Bonnet R, Sutren M, et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and environmental microbiology* 1999; **65**(11): 4799-807.

13. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature* 2012; **486**(7402): 222-7.

14. Wang M, Ahrne S, Jeppsson B, Molin G. Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS microbiology ecology* 2005; **54**(2): 219-31.

15. Tann CJ, Nkurunziza P, Nakakeeto M, et al. Prevalence of bloodstream pathogens is higher in neonatal encephalopathy cases vs. controls using a novel panel of real-time PCR assays. *PLoS One* 2014; **9**(5): e97259.

16. Sabet NS, Subramaniam G, Navaratnam P, Sekaran SD. Simultaneous species identification and detection of methicillin resistance in staphylococci using triplex real-time PCR assay. *Diagn Microbiol Infect Dis* 2006; **56**(1): 13-8.

17. Boutaga K, van Winkelhoff AJ, Vandenbroucke-Grauls CM, Savelkoul PH. Periodontal pathogens: a quantitative comparison of anaerobic culture and real-time PCR. *FEMS immunology and medical microbiology* 2005; **45**(2): 191-9.

18. Sanchez B, Ruiz L, Gueimonde M, Ruas-Madiedo P, Margolles A. Adaptation of bifidobacteria to the gastrointestinal tract and functional consequences. *Pharmacological research : the official journal of the Italian Pharmacological Society* 2013; **69**(1): 127-36.

19. O'Connell Motherway M, Zomer A, Leahy SC, et al. Functional genome analysis of Bifidobacterium breve UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**(27): 11217-22.

20. Ivanov D, Emonet C, Foata F, et al. A serpin from the gut bacterium Bifidobacterium longum inhibits eukaryotic elastase-like serine proteases. *The Journal of biological chemistry* 2006; **281**(25): 17246-52.

21. Schell MA, Karmirantzou M, Snel B, et al. The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract. *Proceedings of the National Academy of Sciences of the United States of America* 2002; **99**(22): 14422-7.

22. Cleusix V, Lacroix C, Dasen G, Leo M, Le Blay G. Comparative study of a new quantitative real-time PCR targeting the xylulose-5-phosphate/fructose-6-phosphate phosphoketolase bifidobacterial gene (xfp) in faecal samples with two fluorescence in situ hybridization methods. *Journal of applied microbiology* 2010; **108**(1): 181-93.

23. Meile L, Rohr LM, Geissmann TA, Herensperger M, Teuber M. Characterization of the D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase gene (xfp) from Bifidobacterium lactis. *Journal of bacteriology* 2001; **183**(9): 2929-36.

24. Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ, Roy NC. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. *Mediators of inflammation* 2013; **2013**: 237921.

25. Altermann E, Buck LB, Cano R, Klaenhammer TR. Identification and phenotypic characterization of the cell-division protein CdpA. *Gene* 2004; **342**(1): 189-97.

26. Azcarate-Peril MA, McAuliffe O, Altermann E, Lick S, Russell WM, Klaenhammer TR. Microarray analysis of a two-component regulatory system involved in acid resistance and proteolytic activity in Lactobacillus acidophilus. *Applied and environmental microbiology* 2005; **71**(10): 5794-804.

27. Buck BL, Azcarate-Peril MA, Klaenhammer TR. Role of autoinducer-2 on the adhesion ability of Lactobacillus acidophilus. *Journal of applied microbiology* 2009; **107**(1): 269-79.

28. Cotter PD, Emerson N, Gahan CG, Hill C. Identification and disruption of lisRK, a genetic locus encoding a two-component signal transduction system involved in stress tolerance and virulence in Listeria monocytogenes. *Journal of bacteriology* 1999; **181**(21): 6840-3.

29. Makarova KS, Koonin EV. Evolutionary genomics of lactic acid bacteria. *Journal of bacteriology* 2007; **189**(4): 1199-208.

30. Xiao Y, Francke C, Abee T, Wells-Bennik MH. Clostridial spore germination versus bacilli: genome mining and current insights. *Food microbiology* 2011; **28**(2): 266-74.

31. Paredes-Sabja D, Setlow P, Sarker MR. Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends in microbiology* 2011; **19**(2): 85-94.

32. Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environmental microbiology* 2012; **14**(11): 2870-90.

33. Andoh A, Fujiyama Y, Hata K, et al. Counter-regulatory effect of sodium butyrate on tumour necrosis factor-alpha (TNF-alpha)-induced complement C3 and factor B biosynthesis in human intestinal epithelial cells. *Clinical and experimental immunology* 1999; **118**(1): 23-9.

34. Duncan SH, Hold GL, Harmsen HJ, Stewart CS, Flint HJ. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. *International journal of systematic and evolutionary microbiology* 2002; **52**(Pt 6): 2141-6.

35. Hamer HM, Jonkers D, Venema K, Vanhoutvin S, Troost FJ, Brummer RJ. Review article: the role of butyrate on colonic function. *Alimentary pharmacology & therapeutics* 2008; **27**(2): 104-19.

36. Louis P, Flint HJ. Development of a semiquantitative degenerate real-time pcr-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. *Applied and environmental microbiology* 2007; **73**(6): 2009-12.

37. Matsuki T, Watanabe K, Fujimoto J, Takada T, Tanaka R. Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces. *Applied and environmental microbiology* 2004; **70**(12): 7220-8.

38. Rinttila T, Kassinen A, Malinen E, Krogius L, Palva A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *Journal of applied microbiology* 2004; **97**(6): 1166-77.

39. Matsuki T, Watanabe K, Fujimoto J, et al. Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Applied and environmental microbiology* 2002; **68**(11): 5445-51.

40. Integrated DNA Technologies. OligoAnalyzer 3.1. 2016. https://eu.idtdna.com/calc/analyzer (accessed 29.6.16 2016).

41. Illumina. Illumina Adapter Sequences. http://support.illumina.com/downloads/illumina-customer-sequence-letter.html (accessed 2016 2016).

42. Salter S, Cox MJ, Turek EM, et al. Reagent contamination can critically impact sequence-based microbiome analyses; 2014.

43. Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl EM. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 2010; **5**(9).

44. Illumina. Indexed sequencing guide. 2016.

45. Mulenga V, Musiime V, Kekitiinwa A, et al. Abacavir, zidovudine, or stavudine as paediatric tablets for African HIV-infected children (CHAPAS-3): an open-label, parallel-group, randomised controlled trial. *Lancet Infect Dis* 2016; **16**(2): 169-79.

46. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. *Archives of disease in childhood* 1995; **73**(1): 17-24.

47. World Health Organization. Global Database on Child Growth and Malnutrition. 2012.

48. Bwakura-Dangarembizi M, Babirye, A., Cook, A., Karungi, C., Kasirye, P., Mapinge, F., Musinguzi, M., Nahirya-Ntege, P., Najjuko, J., Prendergast, A., on behalf of the ARROW trial team. Weight gain in children initiating ART is not increased by nutritional supplements. CROI.

49. Kekitiinwa A, Lee KJ, Walker AS, et al. Differences in factors associated with initial growth, CD4, and viral load responses to ART in HIV-infected children in Kampala, Uganda, and the United Kingdom/Ireland. *J Acquir Immune Defic Syndr* 2008; **49**(4): 384-92.

50. Musiime V, Kayiwa J, Kiconco M, et al. Response to Antiretroviral Therapy of HIV Type 1-Infected Children in Urban and Rural Settings of Uganda. *AIDS research and human retroviruses* 2012.

51. Kekitiinwa A, Cook A, Nathoo K, et al. Routine versus clinically driven laboratory monitoring and first-line antiretroviral therapy strategies in African children with HIV (ARROW): a 5-year open-label randomised factorial trial. *Lancet* 2013; **381**(9875): 1391-403.

52. Penazzato M, Prendergast A, Tierney J, Cotton M, Gibb D. Effectiveness of antiretroviral therapy in HIV-infected children under 2 years of age. *Cochrane database of systematic reviews (Online)* 2012; **7**: CD004772.

53. Violari A, Cotton MF, Gibb DM, et al. Early antiretroviral therapy and mortality among HIV-infected infants. *The New England journal of medicine* 2008; **359**(21): 2233-44.

54. Marchetti G, Bellistri GM, Borghi E, et al. Microbial translocation is associated with sustained failure in CD4+ T-cell reconstitution in HIV-infected patients on long-term highly active antiretroviral therapy. *AIDS* 2008; **22**(15): 2035-8.

55. Estes JD, Harris LD, Klatt NR, et al. Damaged intestinal epithelial integrity linked to microbial translocation in pathogenic simian immunodeficiency virus infections. *PLoS Pathog* 2010; **6**(8).

56. Bowen AC, Mahe A, Hay RJ, et al. The Global Epidemiology of Impetigo: A Systematic Review of the Population Prevalence of Impetigo and Pyoderma. *PLoS One* 2015; **10**(8): e0136789.

57. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced Escherichia coli genomes. *Microb Ecol* 2010; **60**(4): 708-20.

58. Miller JM, Alachi P. Evaluation of new computer-enhanced identification program for microorganisms: adaptation of BioBASE for identification of members of the family Enterobacteriaceae. *J Clin Microbiol* 1996; **34**(1): 179-81.

59. Jiang W, Lederman MM, Hunt P, et al. Plasma levels of bacterial DNA correlate with immune activation and the magnitude of immune restoration in persons with antiretroviral-treated HIV infection. *J Infect Dis* 2009; **199**(8): 1177-85.

60. Sandler NG, Wand H, Roque A, et al. Plasma Levels of Soluble CD14 Independently Predict Mortality in HIV Infection. *J Infect Dis* 2011.

61. Pilakka-Kanthikeel S, Huang S, Fenton T, Borkowsky W, Cunningham CK, Pahwa S. Increased Gut Microbial Translocation in HIV-infected Children Persists in Virologic Responders and Virologic Failures After Antiretroviral Therapy. *Pediatr Infect Dis J* 2012; **31**(6): 583-91.

62. Klase Z, Ortiz A, Deleage C, et al. Dysbiotic bacteria translocate in progressive SIV infection. *Mucosal Immunol* 2015; **8**(5): 1009-20.

63. Chevalier MF, Petitjean G, Dunyach-Remy C, et al. The Th17/Treg ratio, IL-1RA and sCD14 levels in primary HIV infection predict the T-cell activation set point in the absence of systemic microbial translocation. *PLoS Pathog* 2013; **9**(6): e1003453.

64. Kramski M, Gaeguta AJ, Lichtfuss GF, et al. Novel sensitive real-time PCR for quantification of bacterial 16S rRNA genes in plasma of HIV-infected patients as a marker for microbial translocation. *J Clin Microbiol* 2011; **49**(10): 3691-3.

65. Abad-Fernandez M, Vallejo A, Hernandez-Novoa B, et al. Correlation between different methods to measure microbial translocation and its association with immune activation in long-term suppressed HIV-1-infected individuals. *J Acquir Immune Defic Syndr* 2013; **64**(2): 149-53.

66. BenMarzouk-Hidalgo OJ, Torres-Cornejo A, Gutierrez-Valencia A, Ruiz-Valderas R, Viciana P, Lopez-Cortes LF. Differential effects of viremia and microbial translocation on immune activation in HIV-infected patients throughout ritonavir-boosted darunavir monotherapy. *Medicine (Baltimore)* 2015; **94**(17): e781.

67. URI Genomics and Sequencing Center. Calculator for determining the number of copies of a template. 29.1.2004 2004. http://cels.uri.edu/gsc/cndna.html2016).

68. Sandler NG, Douek DC. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. *Nat Rev Microbiol* 2012; **10**(9): 655-66.

69. Wittkop L, Bitard J, Lazaro E, et al. Effect of cytomegalovirus-induced immune response, self antigen-induced immune response, and microbial translocation on chronic immune activation in successfully treated HIV type 1-infected patients: the ANRS CO3 Aquitaine Cohort. *J Infect Dis* 2013; **207**(4): 622-7.

70. Madrid L, Noguera-Julian A, Falcon-Neyra L, et al. Microbial translocation and T cell activation are not associated in chronic HIV-infected children. *AIDS* 2014; **28**(13): 1989-92.

71. Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. *Applied and environmental microbiology* 2000; **66**(4): 1328-33.

72. Ellwood M, Nomura M. Deletion of a ribosomal ribonucleic acid operon in Escherichia coli. *Journal of bacteriology* 1980; **143**(2): 1077-80.

73. National Centre for Biotechnology Information. dnaK chaperone Hsp70, with co-chaperone DnaJ [Escherichia coli str. K-12 substr. MG1655]. Genbank. Genbank: U.S. National Library of Medicine.

74. Schlein RA, Kudlick EM, Reindorf CA, Gregory J, Royal GC. Toothbrushing and transient bacteremia in patients undergoing orthodontic treatment. *Am J Orthod Dentofacial Orthop* 1991; **99**(5): 466-72.

75. Harrison LM, Morris JA, Lauder RM, Telford DR. Staphylococcal pyrogenic toxins in infant urine samples: a possible marker of transient bacteraemia. *J Clin Pathol* 2009; **62**(8): 735-8.

76. Erridge C, Attina T, Spickett CM, Webb DJ. A high-fat meal induces low-grade endotoxemia: evidence of a novel mechanism of postprandial inflammation. *Am J Clin Nutr* 2007; **86**(5): 1286-92.

77. Davenport P, Land KJ. Isolation of Leclercia adecarboxylata from the blood culture of an asymptomatic platelet donor. *Transfusion* 2007; **47**(10): 1816-9.

78. Doyle R. Placental, oral and vaginal microbiomes and birth outcomes in rural Malawi. London: University College London; 2016.

79. National Center for Biotechnology Information. Deinococcus. Taxonomy Browser; 2016. p. Taxonomy ID: 1298.

80. National Centre for Biotechnology Information. Rhodobacter. Taxonomy Browser; 2016. p. Taxonomy ID: 1060.

81. Information. NCfB. Actinomyces. Taxonomy Browser; 2016. p. Taxonomy ID: 1654.

82. Springer J, Loeffler J, Heinz W, et al. Pathogen-specific DNA enrichment does not increase sensitivity of PCR for diagnosis of invasive aspergillosis in neutropenic patients. *J Clin Microbiol* 2011; **49**(4): 1267-73.

83. Hunt PW, Sinclair E, Rodriguez B, et al. Gut epithelial barrier dysfunction and innate immune activation predict mortality in treated HIV infection. *J Infect Dis* 2014; **210**(8): 1228-38.

84. Perkins MR, Bartha I, Timmer JK, et al. The Interplay Between Host Genetic Variation, Viral Replication, and Microbial Translocation in Untreated HIV-Infected Individuals. *J Infect Dis* 2015; **212**(4): 578-84.

85. Olwenyi OA, Naluyima P, Cham F, et al. Brief Report: Differential Associations of Interleukin 6 and Intestinal Fatty Acid-Binding Protein With Progressive Untreated HIV-1 Infection in Rakai, Uganda. *J Acquir Immune Defic Syndr* 2016; **72**(1): 15-20.

86. Tincati C, Merlini E, Braidotti P, et al. Impaired gut junctional complexes feature late-treated individuals with suboptimal CD4+ T-cell recovery upon virologically-suppressive cART. *AIDS* 2016.

87. Cliffe LJ, Humphreys NE, Lane TE, Potten CS, Booth C, Grencis RK. Accelerated intestinal epithelial cell turnover: a new mechanism of parasite expulsion. *Science* 2005; **308**(5727): 1463-5.

88. de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. *Am J Clin Nutr* 1996; **64**(4): 650-8.

89. Kruger HS. Anthropometry and HIV-infected children in Africa. In: Preedy VR, ed. Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease: Springer Science+Business Media, LLC 2012; 2012.

90. MR BL, Drouin O, Bartlett G, et al. Incidence and Prevalence of Opportunistic and Other Infections and the Impact of Antiretroviral Therapy Among HIV-infected Children in Low- and Middle-income Countries: A Systematic Review and Meta-analysis. *Clin Infect Dis* 2016; **62**(12): 1586-94.

91. Balagopal A GL, Franco V, Russell J, Smeaton L, Clements J, Thomas D, Ray S, Gupta A, and ACTG A5175 Team Serum Inhibits Detection of Microbial Translocation in HIV-1 and SIV Infection: ACTG NWCS 319. *18th Conference on Retroviruses and Opportunistic Infections, Boston 27211- 3211* 2011: Abstract 306.

92. Papasavvas E, Azzoni L, Foulkes A, et al. Increased microbial translocation in </= 180 days old perinatally human immunodeficiency virus-positive infants as compared with human immunodeficiency virus-exposed uninfected infants of similar age. *Pediatr Infect Dis J* 2011; **30**(10): 877-82.

93. Wallet MA, Rodriguez CA, Yin L, et al. Microbial translocation induces persistent macrophage activation unrelated to HIV-1 levels or T-cell activation following therapy. *AIDS* 2010; **24**(9): 1281-90.

94. Brenchley JM, Price DA, Schacker TW, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med* 2006; **12**(12): 1365-71.

95. Somsouk M, Estes JD, Deleage C, et al. Gut epithelial barrier and systemic inflammation during chronic HIV infection. *AIDS* 2015; **29**(1): 43-51.

96. Dillon SM, Lee EJ, Kotter CV, et al. An altered intestinal mucosal microbiome in HIV-1 infection is associated with mucosal and systemic immune activation and endotoxemia. *Mucosal Immunol* 2014; **7**(4): 983-94.

97. Monaco CL, Gootenberg DB, Zhao G, et al. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host Microbe* 2016; **19**(3): 311-22.

98. Paquin-Proulx D, Ching C, Vujkovic-Cvijin I, et al. Bacteroides are associated with GALT iNKT cell function and reduction of microbial translocation in HIV-1 infection. *Mucosal Immunol* 2016.

99. Pedersen KK, Manner IW, Seljeflot I, et al. Monocyte activation, but not microbial translocation, is independently associated with markers of endovascular dysfunction in HIV-infected patients receiving cART. *J Acquir Immune Defic Syndr* 2014; **67**(4): 370-4.

100. Redd AD, Dabitao D, Bream JH, et al. Microbial translocation, the innate cytokine response, and HIV-1 disease progression in Africa. *Proceedings of the National Academy of Sciences of the United States of America* 2009; **106**(16): 6718-23.

101. Damien P, Cognasse F, Eyraud MA, et al. LPS stimulation of purified human platelets is partly dependent on plasma soluble CD14 to secrete their main secreted product, soluble-CD40-Ligand. *BMC Immunol* 2015; **16**: 3.

102. Lloyd-Jones KL, Kelly MM, Kubes P. Varying importance of soluble and membrane CD14 in endothelial detection of lipopolysaccharide. *J Immunol* 2008; **181**(2): 1446-53.

103. Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. *Paediatr Int Child Health* 2014; **34**(4): 250-65.

104. Prendergast AJ, Rukobo S, Chasekwa B, et al. Stunting is characterized by chronic inflammation in Zimbabwean infants. *PLoS One* 2014; **9**(2): e86928.

105. Vesterbacka J, Barqasho B, Haggblom A, Nowak P. Effects of Co-Trimoxazole on Microbial Translocation in HIV-1-Infected Patients Initiating Antiretroviral Therapy. *AIDS research and human retroviruses* 2015; **31**(8): 830-6.

106. Merga Y, Campbell BJ, Rhodes JM. Mucosal barrier, bacteria and inflammatory bowel disease: possibilities for therapy. *Dig Dis* 2014; **32**(4): 475-83.

107. Banerjee S, Sindberg G, Wang F, et al. Opioid-induced gut microbial disruption and bile dysregulation leads to gut barrier compromise and sustained systemic inflammation. *Mucosal Immunol* 2016.

108. Grishin A, Bowling J, Bell B, Wang J, Ford HR. Roles of nitric oxide and intestinal microbiota in the pathogenesis of necrotizing enterocolitis. *Journal of pediatric surgery* 2016; **51**(1): 13-7.

109. Wong M, Barqasho B, Ohrmalm L, Tolfvenstam T, Nowak P. Microbial translocation contribute to febrile episodes in adults with chemotherapy-induced neutropenia. *PLoS One* 2013; **8**(7): e68056.

110. Harris KA, Hartley JC. Development of broad-range 16S rDNA PCR for use in the routine diagnostic clinical microbiology service. *J Med Microbiol* 2003; **52**(Pt 8): 685-91.

111. Saglani S, Harris KA, Wallis C, Hartley JC. Empyema: the use of broad range 16S rDNA PCR for pathogen detection. *Archives of disease in childhood* 2005; **90**(1): 70-3.

112. Kenny J. The impact of HIV and Antiretroviral therapy on the cardiovascular system of HIV-infected children. London: University College London; 2016.

113. Hakim J. MV, Szubert A.J., Siika A., Mallewa J., Agutu C., Pett S.L., Bwakura-Dangarembizi M., Lugenwa A., Kaunda S., Karoney M., Maitland K., Griffiths A., Kityo C., Mugyenyi P., Prendergast A. J., Walker A.S., Gibb D.M., and the REALITY trial team Enhanced infection prophylaxis reduces mortality in severely immunosuppressed HIV-infected adults and older children initiating antiretroviral therapy in Kenya, Malawi, Uganda and Zimbabwe: the REALITY trial. World AIDS conference; 2016 21.7.16; Durban, South Africa; 2016. p. Abstract number 10454.

19 Appendices

Appendix A Species in Microbial Mock Communities
Appendix B Tables of barcoded primer sequences for NGS
Appendix C QIIME script settings
Appendix D Missing samples across groups and time points
Appendix E Paper: “Evidence Microbial Translocation Occurs in HIV-Infected Children in the United Kingdom”; AIDS Research and Human Retroviruses; October 2013
Appendix F Conference Abstract “Microbial translocation does not drive immune activation in Ugandan children with HIV”; Conference for Retroviruses and Opportunistic Infections February 2017, Seattle, Washington; Abstract number 188

19.1 Appendix A Species in Microbial Mock Communities

Organism	NCBI Reference Sequence
Acinetobacter baumannii, strain 5377	NC_009085
Actinomyces odontolyticus, strain 1A.21	NZ_AAYI02000000
Bacillus cereus, strain NRS 248	NC_003909
Bacteroides vulgatus, strain ATCC® 8482™	NC_009614
Clostridium beijerinckii, strain NCIMB 8052	NC_009617
Deinococcus radiodurans, strain R1 (smooth)	NC_001263, NC_001264
Enterococcus faecalis, strain OG1RF	NC_17316
Escherichia coli, strain K12, substrain MG1655	NC_000913
Helicobacter pylori, strain 26695	NC_000915
Lactobacillus gasseri, strain 63 AM	NC_008530
Listeria monocytogenes, strain EGDe	NC_003210
Neisseria meningitidis, strain MC58	NC_003112
Propionibacterium acnes, strain KPA171202	NC_006085
Pseudomonas aeruginosa, strain PAO1-LAC	NC_002516
Rhodobacter sphaeroides, strain ATH 2.4.1	NC_007493, NC_007494
Staphylococcus aureus, strain TCH1516	NC_010079
Staphylococcus epidermidis, FDA strain PCI 1200	NC_004461
Streptococcus agalactiae, strain 2603 V/R	NC_004116
Streptococcus mutans, strain UA159	NC_004350
Streptococcus pneumoniae, strain TIGR4	NC_003028

19.2 Appendix B Tables of barcoded primer sequences for NGS

Table 39 Initial Set of Barcoded Primers

785F read 1 primer sequence (Forward primer)

Illumina P5 adapter	Index	Padding	785F primer	Primer ID
AATGATACGGCGACCACCGAGATCTACAC	TCCCTTGTCTCC	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc0
AATGATACGGCGACCACCGAGATCTACAC	ACGAGACTGATT	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc1
AATGATACGGCGACCACCGAGATCTACAC	GCTGTACGGATT	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc2
AATGATACGGCGACCACCGAGATCTACAC	ATCACCAGGTGT	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc3
AATGATACGGCGACCACCGAGATCTACAC	TGGTCAACGATA	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc4
AATGATACGGCGACCACCGAGATCTACAC	ATCGCACAGTAA	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc5
AATGATACGGCGACCACCGAGATCTACAC	GTCGTGTAGCCT	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc6
AATGATACGGCGACCACCGAGATCTACAC	AGCGGAGGTTAG	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc7
AATGATACGGCGACCACCGAGATCTACAC	ATCCTTTGGTTC	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc8
AATGATACGGCGACCACCGAGATCTACAC	TACAGCGCATAC	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc9
AATGATACGGCGACCACCGAGATCTACAC	ACCGGTATGTAC	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc10
AATGATACGGCGACCACCGAGATCTACAC	AATTGTGTCGGA	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc11
AATGATACGGCGACCACCGAGATCTACAC	TGCATACACTGG	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc12
AATGATACGGCGACCACCGAGATCTACAC	AGTCGAACGAGG	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc13
AATGATACGGCGACCACCGAGATCTACAC	ACCAGTGACTCA	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc14
AATGATACGGCGACCACCGAGATCTACAC	GAATACCAAGTC	TACCGGGACTTA	GGATTAGATACCCBRGTAGTC	Fbc15

1175R read 2 primer sequence (Reverse primer)

Illumina P7 adapter	Index	Padding	1175R primer	Primer ID
CAAGCAGAAGACGGCATACGAGAT	GCATATGCACTG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc32
CAAGCAGAAGACGGCATACGAGAT	CAACTCCCGTGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc33
CAAGCAGAAGACGGCATACGAGAT	TTGCGTTAGCAG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc34
CAAGCAGAAGACGGCATACGAGAT	TACGAGCCCTAA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc35
CAAGCAGAAGACGGCATACGAGAT	CACTACGCTAGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc36

CAAGCAGAAGACGGCATACGAGAT	TGCAGTCCTCGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc37
CAAGCAGAAGACGGCATACGAGAT	ACCATAGCTCCG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc38
CAAGCAGAAGACGGCATACGAGAT	TCGACATCTCTT	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc39
CAAGCAGAAGACGGCATACGAGAT	GAACACTTTGGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc40
CAAGCAGAAGACGGCATACGAGAT	GAGCCATCTGTA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc41
CAAGCAGAAGACGGCATACGAGAT	TTGGGTACACGT	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc42
CAAGCAGAAGACGGCATACGAGAT	AAGGCGCTCCTT	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc43
CAAGCAGAAGACGGCATACGAGAT	TAATACGGATCG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc44
CAAGCAGAAGACGGCATACGAGAT	TCGGAATTAGAC	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc45
CAAGCAGAAGACGGCATACGAGAT	TGTGAATTCGGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc46
CAAGCAGAAGACGGCATACGAGAT	CATTCGTGGCGT	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc47
CAAGCAGAAGACGGCATACGAGAT	TACTACGTGGCC	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc48
CAAGCAGAAGACGGCATACGAGAT	GGCCAGTTCCTA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc49
CAAGCAGAAGACGGCATACGAGAT	GATGTTCGCTAG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc50
CAAGCAGAAGACGGCATACGAGAT	CTATCTCCTGTC	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc51
CAAGCAGAAGACGGCATACGAGAT	ACTCACAGGAAT	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc52
CAAGCAGAAGACGGCATACGAGAT	ATGATGAGCCTC	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc53
CAAGCAGAAGACGGCATACGAGAT	GTCGACAGAGGA	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc54
CAAGCAGAAGACGGCATACGAGAT	TGTCGCAAATAG	AACACGTTTTA	ACGTCRTCCCCDCCTTCCTC	Rrcbc55

Table 40 Second Set of Barcoded Primers

785F read 1 primer sequence

Illumina P5 adapter	Index	Padding	785F primer	Primer ID
AATGATACGGCGACCACCGAGATCTACAC	CTCTCTAT	ACGTGCGTACGT	GGATTAGATACCCBRGTAGTC	Fbc0
AATGATACGGCGACCACCGAGATCTACAC	TATCCTCT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc1
AATGATACGGCGACCACCGAGATCTACAC	GTAAGGAG	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc2
AATGATACGGCGACCACCGAGATCTACAC	ACTGCATA	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc3
AATGATACGGCGACCACCGAGATCTACAC	AAGGAGTA	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc4
AATGATACGGCGACCACCGAGATCTACAC	CTAAGCCT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc5
AATGATACGGCGACCACCGAGATCTACAC	CGTCTAAT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc6
AATGATACGGCGACCACCGAGATCTACAC	TCTCTCCG	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc7
AATGATACGGCGACCACCGAGATCTACAC	TCGACTAG	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc8
AATGATACGGCGACCACCGAGATCTACAC	TTCTAGCT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc9
AATGATACGGCGACCACCGAGATCTACAC	CCTAGAGT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc10
AATGATACGGCGACCACCGAGATCTACAC	GCGTAAGA	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc11
AATGATACGGCGACCACCGAGATCTACAC	CTATTAAG	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc12
AATGATACGGCGACCACCGAGATCTACAC	AAGGCTAT	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc13
AATGATACGGCGACCACCGAGATCTACAC	GAGCCTTA	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc14
AATGATACGGCGACCACCGAGATCTACAC	TTATGCGA	ACGTACGTACGT	GGATTAGATACCCBRGTAGTC	Fbc15

1175R read 2 primer sequence

Illumina P7 adapter	Index	Padding	1175R primer	Primer ID
CAAGCAGAAGACGGCATACGAGAT	TAAGGCGA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc32
CAAGCAGAAGACGGCATACGAGAT	CGTACTAG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc33
CAAGCAGAAGACGGCATACGAGAT	AGGCAGAA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc34
CAAGCAGAAGACGGCATACGAGAT	TCCTGAGC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc35
CAAGCAGAAGACGGCATACGAGAT	GGACTCCT	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc36
CAAGCAGAAGACGGCATACGAGAT	TAGGCATG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc37
CAAGCAGAAGACGGCATACGAGAT	CTCTCTAC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc38
CAAGCAGAAGACGGCATACGAGAT	CGAGGCTG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc39
CAAGCAGAAGACGGCATACGAGAT	AAGAGGCA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc40
CAAGCAGAAGACGGCATACGAGAT	GTAGAGGA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc41
CAAGCAGAAGACGGCATACGAGAT	GCTCATGA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc42
CAAGCAGAAGACGGCATACGAGAT	ATCTCAGG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc43
CAAGCAGAAGACGGCATACGAGAT	ACTCGCTA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc44
CAAGCAGAAGACGGCATACGAGAT	GGAGCTAC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc45
CAAGCAGAAGACGGCATACGAGAT	GCGTAGTA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc46
CAAGCAGAAGACGGCATACGAGAT	CGGAGCCT	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc47
CAAGCAGAAGACGGCATACGAGAT	TACGCTGC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc48
CAAGCAGAAGACGGCATACGAGAT	ATGCGCAG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc49
CAAGCAGAAGACGGCATACGAGAT	TAGCGCTC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc50
CAAGCAGAAGACGGCATACGAGAT	ACTGAGCG	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc51
CAAGCAGAAGACGGCATACGAGAT	CCTAAGAC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc52
CAAGCAGAAGACGGCATACGAGAT	CGATCAGT	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc53
CAAGCAGAAGACGGCATACGAGAT	TGCAGCTA	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc54
CAAGCAGAAGACGGCATACGAGAT	TCGACGTC	AGTCAGTCAGCC	ACGTCRTCCCCDCCTTCCTC	Rrcbc55

19.3 Appendix C QIIME script settings

Assemble paired end reads using FLASH

#This script assembles paired end reads

./flash -m 110 -M 150 -d /home/qiime/felicity/
/home/qiime/felicity/Undetermined_S0_L001_R1_001.fastq
/home/qiime/felicity/Undetermined_S0_L001_R2_001.fastq

cd /home/qiime/

#This script is to reverse the complementarity of the index read. Run twice to extract both indexes before combining to one barcode matching mapping file.

python extract_bcs_from_fastq.py felicity/1_S1_L001_I1_001.fastq
felicity/extracted_barcodes1.fastq felicity/extracted_reads1.fastq 12 True

python extract_bcs_from_fastq.py felicity/1_S1_L001_I2_001.fastq
felicity/extracted_barcodes2.fastq felicity/extracted_reads2.fastq 12 False

python combine_fastq_barcodes.py felicity/extracted_barcodes1.fastq
felicity/extracted_barcodes2.fastq felicity/combined_barcodes.fastq

#Extra script in FLASH directory that writes barcodes in same order as assembler. Assembler works in random parallel order leading to random list of reads in output fastq file. Need this script to make sure both read file and barcode file headers match.

python remove_unused_barcodes.py /home/qiime/felicity/combined_barcodes.fastq
/home/qiime/felicity/out.extendedFrags.fastq
/home/qiime/felicity/updated.barcodes.outfile.fastq

#Split libraries and demultiplex illumina fastq file (matching barcoded amplicons to samples in mapping file). Use --barcode_type as argument for how long barcode is- in this case 24.

split_libraries_fastq.py -i /home/qiime/felicity/out.extendedFrags.fastq -o
/home/qiime/felicity/split_lib/ -b
/home/qiime/felicity/updated.barcodes.outfile.fastq -m
/home/qiime/felicity/Miseq_2_mapping.txt --barcode_type 24

```
#pick OTUS reference

parallel_pick_otus_uclust_ref.py -i felicity/split_lib/seqs.fna -o felicity/pick_otus/ -r
qiime_software/gg_otus-13_8-release/rep_set/97_otus.fasta -O 8 -T


pick_rep_set.py -i pick_otus/seqs_otus.txt -f pick_otus/seqs.fna -m most_abundant -
o rep_set.fna


#Parallel align sequences using PyNAST

parallel_align_seqs_pynast.py -i rep_set.fna -o pynast_aligned_seqs/ -T


#Filter alignment against lanemask file

filter_alignment.py -o pynast_aligned_seqs/ -i
pynast_aligned_seqs/rep_set_aligned.fasta


#Build phylogenetic tree

make_phylogeny.py -i pynast_aligned_seqs/rep_set_aligned_pfiltered.fasta -o
rep_set.tre


#Assign Taxonomy via uclust


parallel_assign_taxonomy_uclust.py -i rep_set.fna -o
pick_otus/uclust_assigned_taxonomy/ -O 2


#Build OTU table


make_otu_table.py -i pick_otus/seqs_otus.txt -t
pick_otus/uclust_assigned_taxonomy/rep_set_tax_assignments.txt -o
otu_table.biom


#Print OTU table summary to text file


biom summarize-table -i otu_table.biom -o pick_otus/otu_table_summary.txt



#Print OTU table to text file
```

```
biom convert -i otu_table.biom -o pick_otus/otu_table.txt -b --header-key taxonomy


#Filter out OTUs that appear less than 0.001


filter_otus_from_otu_table.py -i otu_table.biom -o otu_table_mcf0001.biom --
min_count_fraction 0.001



#Print filtered OTU table to text file


biom convert -i otu_table_mcf0001.biom -o otu_table_mcf0001.txt -b --header-key
taxonomy


#Print OTU table summary to text file


biom summarize-table -i /home/qiime/felicity/pick_otus/otu_table_mcf0001.biom -o
/home/qiime/felicity/pick_otus/otu_table_mcf0001_summary.txt


#Perform a single random rarefaction on otu table to produce even sampling depth
across all samples (In this case 1000 sequences per sample)


single_rarefaction.py -i /home/qiime/felicity/pick_otus/otu_table_mcf0001.biom -o
/home/qiime/felicity/pick_otus/otu_table_mcf0001_1000.biom -d 1
```

19.4 Appendix D Missing samples across groups and time points

Baseline	qPCR					SYBR					I-FABP				
	Naïve	Experienced	Gulu	Naïve Controls	Experienced Controls	Naïve	Experienced	Gulu	Naïve Controls	Experienced Controls	Naïve	Experienced	Gulu	Naive Controls	Experienced Controls
total	119	22	55	89	20	119	22	55	89	20	119	22	55	89	20
total missing	9	0	4	0	0	9	0	4	0	0	10	2	9	0	0
Sample not available	9		4			9		4			9		8		
Assay Failure											1	2	1		
Unknown															
Week 12	**qPCR**			**SYBR**			**I-FABP**								
	Naïve	Experienced	Gulu	Naïve	Experienced	Gulu	Naïve	Experienced	Gulu						
Died	0	0	2	0	0	2	0	0	2						
Loss to Follow up	0	0	0	0	0	0	0	0	0						
total excluding deaths and loss to follow up by W12	119	22	53	119	22	53	119	22	53						
total missing	11	3	5	11	3	6	12	3	16						
Sample not available	11	3	5	11	3	5	11	3	15						
Assay Failure				0	0	1	1		1						
Unknown															
Week 72	**qPCR**			**SYBR**			**I-FABP**								
	Naïve	Experienced	Gulu	Naïve	Experienced	Gulu	Naïve	Experienced	Gulu						
Died	3	0	2	3	0	2	3	0	2						
Loss to Follow up	4	0	2	4	0	2	4	0	2						
Unknown	4	1													
total excluding deaths and loss to follow up by W72	112	22	51	112	22	51	112	22	51						
total missing	2	0	7	3	0	7	3	1	12						
Sample not available	2		7	3		7	3		10						
Assay Failure								1	2						

19.5 Appendix E Paper: “Evidence Microbial Translocation Occurs in HIV-Infected Children in the United Kingdom”; AIDS Research and Human Retroviruses; October 2013

19.6 Appendix F Conference Abstract "Microbial translocation does not drive immune activation in Ugandan children with HIV"; Conference for Retroviruses and Opportunistic Infections February 2017, Seattle, Washington; Abstract number 188

Microbial translocation does not drive immune activation in Ugandan children with HIV

Fitzgerald F., Lhomme E., Harris K., Kenny J., Doyle R., Kityo C., Walker A.S., Thiebaut R., Klein N. and the CHAPAS3 Trial Team

FF, JK, NK: Infection, Immunity and Inflammation Programme, UCL Great Ormond Street Institute of Child Health, London, UK

EL, RT: INSERM U1219, INRIA SISTM, Bordeaux Univ., Bordeaux, France

KH : Microbiology, Camelia Botnar laboratories, Great Ormond Street NHS Foundation Trust, London, UK

RD: Division of Infection and Immunity, UCL, London, UK

CK : Joint Clinical Research Centre, Kampala, Uganda

ASW : MRC Clinical Trials Unit at UCL, London, UK

Background

Immune activation, potentially driven by microbial translocation (MT), is linked to increased morbidity and mortality despite ART in HIV infection. This study aimed to investigate MT as a driver of poor outcome in HIV-infected African children

Methods

The study population comprised ART-naïve and ART-experienced children recruited to Ugandan sites of the CHAPAS-3 Trial (ISRCTN69078957), a toxicity trial of fixed-dose combination ART regimens, and HIV-uninfected age-matched controls from the same communities. HIV-infected children were followed up for 96 weeks including viral suppression, immune recovery and anthropometric measures. MT was assessed using a panel of specific bacterial polymerase chain reactions (PCRs), broad-range 16S rDNA PCR and next generation sequencing (NGS)(Illumina® method). 19 markers of immune activation were measured including cellular and humoral markers. Intestinal fatty acid binding protein (I-FABP) was used to quantify gut damage. Analysis including cluster analysis was performed in R Studio (v.0.99.896).

Results

250 children were included: 119 ART naïve (median age 2.8 years, interquartile range (IQR) 1.7-4, median baseline CD4% 20, IQR 14-24) and 22 ART experienced children

(median age 6.5 years IQR 5.9-9.2, median baseline CD4% 34, IQR 31-39) and 109 age-matched HIV-uninfected controls. Immune recovery was good in both HIV-infected groups, and viral load suppression <100 copies/ml was achieved in 76% (ART-naïve) and 91% (ART-experienced) at 96 weeks. Four children died (all ART-naïve). Specific PCRs for *Bifidobacterium, Lactobacillus, Fusobacterium,* and *Streptoccocus pyogenes* were negative in all groups at most time points. *Enterobacteriaceae* PCR was positive in 23-55% of samples at different time points and was higher in controls than ART-naive at baseline (p=0.03). There was no significant change in proportions positive over time. Baseline *Staphylococcus aureus* proportions positive were 23% (95% CI 8-45%) in ART-experienced and 10% (95% CI 5-16%) in ART-naïve (no difference vs controls, p=0.22 & 0.81 for age matched naïve/experienced) and decreased non-significantly to 4.8% (95% CI 1-23%) vs 6.7% (95% CI 2.5-12.4%) at week 72. At baseline using NGS, very low levels of microbial DNA were found in both HIV-infected groups, including *S.aureus, Enterobacteriaceae, Veillonellae* & *Clostridiales*. Cluster analysis showed… The improvement in immune activation markers over time had no relationship to markers of MT.

Conclusion

Levels of bacterial DNA were low in all children regardless of HIV or ART status. Immune activation decreased over time on ART. MT may not be a significant driver of immune activation in this setting.

Table 1. Characteristics of the clustering and distribution of microbial translocation markers among children from a Ugandan site of CHAPAS 3 trial.

Characteristics	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**P-value**
N (%)	105 (42)	120 (48)	11 (4)	13 (5)	
ART at baseline					0.0036
ART naive	57 (54)	50 (42)	6 (55)	7 (54)	
ART experienced	14 (13)	5 (4)	0 (0)	3 (23)	
HIV negative	34 (32)	65 (54)	5 (45)	3 (23)	
Virally suppressed at W96 (<100 copies/ml)					<0.0001
Yes	65 (62)	32 (27)	3 (27)	3 (23)	
No	6 (6)	14 (12)	3 (27)	6 (46)	
HIV negative	34 (32)	65 (54)	5 (45)	3 (23)	
Viral load missing	0 (0)	9 (8)	0 (0)	1 (8)	
I-FABP pg/ml*	176 (139)	187 (191)	158 (72)	192 (115)	0.93
16S rDNA **	48 (46)	61 (51)	6 (55)	6 (46)	0.86
Bifidobacterium spp. **	0 (0)	0 (0)	0 (0)	0 (0)	-
*Staphylococcus aureus***	3 (4)	5 (10)	0 (0)	0 (0)	0.58
*Streptococcus pyogenes***	0 (0)	0 (0)	1 (9)	0 (0)	0.05
*Fusobacterium spp.***	1 (1)	1 (1)	1 (9)	0 (0)	0.21
*Enterobacteriaceae***	47 (45)	30 (26)	5 (45)	1 (8)	0.006
Staphylococcus spp. **	0 (0)	0 (0)	0 (0)	0 (0)	-
*Lactobacillus spp.***	1 (1)	1 (1)	0 (0)	0 (0)	0.54

*I-FABP: Intestinal fatty acid binding protein, mean (standard deviation)

**Number of samples testing positive using PCR. Sensitivities from 0.5-500 colony forming unit (CFU) equivalents when compared with standard of known CFUs depending on assay.